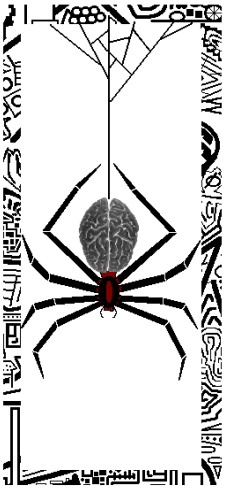
THINKING Outside the BOX

A THEORY OF EMBODIED AND EMBEDDED CONCEPTS



Marco van Leeuwen

The research presented in this doctoral thesis was carried out at the Department of Philosophy of Radboud University Nijmegen, The Netherlands.

ISBN/EAN: 978-90-9024487-7

Cover design by Marco van Leeuwen

Printed by Universal Press, Veenendaal, The Netherlands

Thinking Outside the Box:
A Theory of Embodied and Embedded Concepts

Een wetenschappelijke proeve op het gebied van de Filosofie

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag 22 september 2009
om 15.30 uur precies

door

Marco van Leeuwen
geboren op 25 januari 1976
te Schiedam

Promotor:

Prof. dr. M.V.P. Slors

Manuscriptcommissie:

Prof. dr. R.A. van der Sandt, voorzitter
Prof. dr. A.M.T. Bosman
Prof. dr. E. Myin (Universiteit Antwerpen)

Table of Contents **Page**

[10 - Evaluation, Application and Conclusion]

[1 - Introduction]

1.1 - Exorcising Cartesianism

In modern philosophy of cognition, the Cartesian schema has long been the dominant way of characterizing the role of the mind. The notion of a mind that is, at least in principle, separable from the body because it is an entirely different kind of entity, is firmly embedded in everyday parlance, but even in philosophical and psychological theories that are proclaimed to have done away with Cartesianism, some rudiments of the old schema sometimes remain.

These Cartesian rudiments often take the form of what Ryle (1949) called a 'category mistake', in which a term from one logico-linguistic category is incorrectly applied to something that would require the application of a term from a wholly different category. An example of an error of this kind would be a case in which a capacity or activity of the agent *as a whole* is somehow attributed to something 'inside the head', be it a brain region or a particular functional state. Bennett and Hacker (2003), for instance, provide a lengthy critique of such cases. In their book, neuroscientific models in which the brain (or a specific brain region) is claimed to 'see' and 'hear' most obviously fall prey to the charge of hidden Cartesianism, but few today hold such views explicitly. However, even more subtle formulations involving, for instance, neural correlates of aspects of visual scenes being identified as *representations* of those aspects that need to be processed neurally, are picked apart and criticized as harbouring old Cartesian rudiments in a more or less implicit fashion. Whatever can be said about the validity of the arguments developed by Bennet and Hacker[NOTE 1], their inclination towards doing away with these old theoretical impediments is shared by many others, some of whom will be mentioned below.

The main bulwark of Cartesianism in the modern era is *cognitivism*, which can be characterised by mentioning three central hypotheses: (1) representations, internal states that stand in for or symbolise external states, are the constituents of mental phenomena; (2) the syntaxis (form) rather than semantic content of these representations is most significant; (3) it is possible to specify the rules that govern the form-based transformations of representations into other representations. The most important Cartesian aspect of cognitivism involves the idea that cognition is to be thought of as a kind of symbol manipulation that takes place somewhere in the brain, irrespective of the precise physical instantiation of these symbols.

This is the basic picture that underlies the computer-metaphor of the mind: cognition is, in essence, the manipulation of language-like symbols governed by explicit rules. One of the main proponents of this kind of thinking about the mind is Jerry Fodor (1975).

In recent years, the discontent with the Cartesian and/or cognitivistic picture of the relationship between mind and body - and in fact the idea that there is

to be talk of a *relationship* at all - has grown considerably, coinciding with and feeding off the groundswell of theories proclaiming that mind and body are in actuality an inseparable unity. These theories of *embodied* cognition state that explaining the mind requires taking into account the way in which that mind controls the body, and how the properties of that body in turn enable and/or constrain the activities of the mind. Expanding upon this notion, many maintain that cognition is also *embedded*, meaning that properties of the environment (i.e. factors external to the organism) are relevant to the explanation of cognitive processes as well. In a particular theory of embodied and embedded cognition, these two factors might be attributed varying weight, their most radical implementation yielding the idea that cognition *is* behaviour (almost resembling 'good old' philosophical behaviourism - Thelen et al. (2001) defend a theory - to be discussed in section 3.2 - that is like this, in a way, attempting to abolish any and all need for talk of representations), or that external processes form an integral component of cognitive processes (Clark's [1997] scaffolding, McCulloch's [2003] phenomenal externalism).

This variance in the weights of theory components contributes to the fact that it is not always clear what 'embodiment' and 'embeddedness' are supposed to mean (Ziemke, 2003). In reality, 'embodied and embedded cognition' is not so much a coherent theory as it is a collection of closely (and sometimes not so closely) related approaches to explaining cognition.

For any particular flavour of theory about embodied/embedded cognition, the kinds of problems that are addressed, the contributing disciplines, as well as the kinds of theories viewed as inspirational might all be selected from a rather wide range of possibilities. This adds to the impression that 'the embodied/embedded cognition paradigm' constitutes a grab-bag of approaches rather than a fully realised, consistent and coherent theory: this 'paradigm' lies at a nexus of neuroscience, cognitive psychology, dynamical systems theory and philosophy of cognition. In it, we see approaches that are anti-Cartesian, anti-computationalist (Clark, 1997), pro-ecological, often pro-J.J.Gibson (see Gibson, 1979; Varela et al., 1991), and dynamical (Port and Van Gelder, 1995; Thelen et al., 2001). We see attempts to integrate knowledge from phenomenological traditions, often Merleau-Ponty (Thompson, 2007) and/or Heidegger (Clark, 1997), and sometimes inspiration is sought from Aristotle (Juarrero, 1999), possibly the pre-Socratics, and even Buddhism (Varela et al. 1991).

1.2 - Varieties of $E_{(i)}C$

I would suggest that part of the reason why the 'embodied and embedded cognition'-paradigm is a somewhat muddled and fractured field of research, lies in the fact that there are several elements in addition to 'embodiment' and 'embeddedness' which can be included in this kind of approach to characterising the mind and cognition. Which elements should be adhered to - and in which way - can vary considerably, depending on who you ask, yet they can all be grouped in roughly the same 'subsection' of the field of

theories about the mind. Quite conveniently - or perhaps confusingly, depending on your inclination - each of these elements can be referred to by a term starting with the letter 'e'. I propose the notation '$E_{(i)}C$' as a more inclusive and neutral way to refer to what, so far, has been dubbed the 'embodied and embedded cognition'-approach, with 'i' as a placeholder for any combination of the theory-components listed below.

Varieties of $E_{(i)}C$

-*Embodied* Cognition
Notation: $E_{(B)}C$

'Embodiment' might take on a meaning as deflated as mere perceptual grounding (Gärdenfors 2000, pg 160-161, where he quotes Jackendoff 1983): what we think and feel is informed by perceptual input, hence whatever we mean by our expressions has, at some point, been run through a throng of perceptual filters, i.e. has been influenced by the way our body (including the perceptual subsystems) functions. A more substantial conception of 'embodiment' can involve claims such as those made by Damasio (1999): processes involved in realizing basic bodily awareness, proprioception and emotional responses are also constitutive of the processes that realise cognition, crucially including *off-line* cognitive processing. Jesse Prinz' (2002) concept empiricism (see section 10.3) is a theory which falls in this general category: according to Prinz, the properties of the body's sensory organs and the properties of concepts are intimately linked.

-*Embedded* Cognition
Notation[NOTE 2]: $E_{(S)}C$

'Embeddedness' can refer to a lot of things, but the main idea is that an agent is to be understood in relation to his environment. One way to flesh out this idea is by referring to J.J. Gibson's (1979) 'affordances', where the properties of (some relevant part of) the environment and the capabilities of the agent conspire to define a range of possibilities for action. Some of the theoretical flavours that characterise this range border on enactive cognition (see below), when the way in which the agent *acts* in this environment is taken into account; other varieties blend into extended cognition (also below), as the definition of what constitutes a mind is modified in such a way that the mind's boundaries extend beyond the agent's skin, and into what would count as 'the environment' in other theories.

-*Extended* Cognition
Notation: $E_{(X)}C$

Claiming cognition, or the mind as such, is 'extended', is to say that at least some mental processes utilise extradermal artifacts and processes in such a way that they should be claimed to form a proper part of the mental proceedings. For Clark's (1997) notion 'scaffolding', which refers to activities

at a fairly innocuous end of this spectrum - e.g. using bits of paper to scribble on while performing calculations, which would mean that these scribblings *support* mental activity -, a case can be made to classify it under the 'embedded' header. However, the stronger claim of objects actually becoming part of an agent's cognitive processes is also possible (Clark and Chalmers 1998): imagine a chronic amnesiac who uses a notepad the way 'normal' people use their memory. In such cases, the mind is said to 'leak out of the skull', and into the environment. For some supporters of this position (e.g. Clark and Chalmers), this is less of a metaphor than one might think.

-*Enactive* Cognition[NOTE 3]
Notation: $E_{(A)}C$

The central tenet of 'enactivism' (Varela et al. 1991) is that cognition should be understood in terms of an interaction-process of body and world. An agent does not *have* a belief in the same way he can have blue eyes or curly hair; rather, having a belief means acting out whatever this belief implies in a minded interplay with the world. This principle applies to sensorimotor activity in particular: seeing, for instance, is not a passive information-processing procedure, but it is a specific mode of interaction with the environment (O'Regan and Noë 2001, Noë 2004). An important driving force behind this view is the explicit notion that representation cannot and should not be invoked to explain most (and some would say *all*) forms of cognition-involving action.

Some dynamicist approaches (i.e. DST-C, the application of dynamical systems theory to cognition) might be characterised as supporting enaction: the distinctive claim of Port and Van Gelder (1995), for instance, is that the study of cognition is essentially about the kind of activity an agent exhibits over time. That is, an agent has a history, and any cognitive process that he might exhibit is crucially interwoven with what he does, how he does it, and also with the ways in which these acts are shaped by and leave their traces on his environment.

An additional rider that might be added to the 'enactivist' component is that action can be *meaningful* to a particular agent, in a particular niche: colour vision - used for detecting prey or food, for instance - can be understood in terms of an agent's interaction with certain features of his niche, and this interaction dynamic is the way it is because of practical, meaningful, evolutionarily significant reasons (Thompson 1995). As such, enactivism is often closely related to the theories of J.J. Gibson (1979) (see also section 8.2).

-*Encultured* Cognition[NOTE 4]
Notation: $E_{(C)}C$

An agent can also be 'encultured', referring to the immersion in a particular socio-cultural context. Such a context has its own rules and regulations that

set it apart from the less layered, less abstract and less symbol-centered (but certainly no less dynamic) *ecological* context covered by the 'embedded' tag. This immersion is the kind of agent-to-world-dynamic studied by cultural anthropologists, and similar to what sociologists have dubbed 'socialization': it is that part of the agent's environment chiefly formed by other people, their behaviour, and the meaningful symbols (speech, text, complex signs and signals) they create. An important regulating dynamic by which the agent can cohabitate with other agents involves learning how to interpret these cues, and the continuing adaptation to new variations on these themes.

Different combinations of these $E_{(\ i\ )}C$-components are possible. Furthermore, each of the components above allows some variance in the kind of ontology it prescribes or implies regarding the mind. For instance, Clark, Chalmers, Gärdenfors and others who pledge adhesion to at least some of these components, subscribe to fairly robust ontological commitments about the mind, whereas J.J. Gibson and dynamicists such as Thelen promote a somewhat 'deflationary' view of cognition, abolishing as much internal processing as possible, and pushing what is left to the agent's sensorimotor periphery.

Another example: some of the classic Continental philosophers - e.g. Merleau-Ponty - might be described as adhering to $E_{(B,\ S,\ A)}C$, stressing, as they do, the primacy of the body, environment and action; the same goes for J.J. Gibson. However, a case can be made for the claim that despite these decidedly non-trivial similarities, Gibson subscribes to notions about representation and phenomenology that differ from those held by Merleau-Ponty. That is, Gibson exorcises them, whereas Merleau-Ponty, despite his anti-cognitivist (i.e. anti-Cartesian) leanings, still operates within an opposition between the conscious subject and the perceived object, at least in his earlier writings.

All signs point to the conclusion that the philosophers and scientists working on theories of embodied/embedded cognition are still in the middle of founding this new paradigm, and with it they intend to assimilate the best of a number of relevant philosophical and psychological traditions, while discarding those elements that have proven to be ineffective. The one thing that binds together this multitextured patchwork of approaches, is the aim to offer an alternative to the classic theories of cognition, Cartesian dualism and computationalism in particular. So it might make sense to speak of the 'embodied/embedded cognition'-*inclination* (rather than paradigm), this inclination resulting in a somewhat non-homogeneous array of research programmes, bound together by a common 'enemy'[NOTE 5]. Here, some clean-up and structure is needed to streamline the efforts in this burgeoning field.

1.3 - Central Question

To provide some of that structure is the purpose of this book. The main focus of this book involves a proposal to help theories of embodied/embedded cognition account for actual *thinking* - cognition, by providing the first piece of that puzzle: a theory of concepts. There is a respectable number of successful embodied/embedded models and experiments involving rather basic cognitive or proto-cognitive processes, but few of them manage to provide the tools with which to understand 'higher' forms of cognition in an embodied and embedded context.

Clark (1997), for instance, introduces the cockroach as the new paradigmatic example of cognitive behaviour. He uses the distributed character of the mechanisms responsible for controlling the cockroach's actions to underscore the idea that the brain should be conceptualised as an *embodied controller*, thereby replacing the old computer-metaphor of the mind prevalent in computationalism (including the idea of mental operations as centralised symbol processing). Clark's coackroach-example (as well as the majority of the other examples, anecdotes and vignettes he uses) yields a very effective heuristic tool, and it certainly helps redefine the kinds of theories that are needed to account for basic sensorimotor interaction with the world, but it says little about the kind of cognition involved in thinking, writing a novel, composing music, playing chess and any of a million other activities involving higher cognition - even having a discussion with a colleague can be understood in a way that involves, but is quite definitely not exhausted by, basic embodied processes. I am certain that Clark knows this, as he discusses the useful notion 'representation-hungry problems' (see also section 7.4), but the impression remains that, after Clark has said and done all he planned to, there is still more to be said about the more highly developed activities of the mind.

Similar remarks can be made about the field of dynamical systems theory as applied to cognition. Thelen et al. (2001) build a solid case in support of a theory of (some aspects of) cognition that does not require internal representation (of the computationalist/cognitivist kind). However, their model does not address higher cognition[NOTE 6]. As such, the model *does* constitute an important first step towards a dynamicist theory of cognition, but it does not yet address cognition as we understand it in everyday parlance. Of course, part of the 'embodied/embedded'-project is geared towards changing exactly this outdated paradigm of cognition, and psychology and the philosophy of cognition are likely to become healthier disciplines because of such changes, but this new paradigm does not offer a snug fit: there is quite a bit of room left in the mind (metaphorically speaking, obviously) that is not covered by this new vernacular.

Another example: Bermudez (1998) addresses the role of basic sensorimotor processes for self-consciousness, but these processes are all held to have their greatest influence at the sub- or non-conceptual level. Certainly nonconceptual content (inasmuch as there can be talk of 'content'

here, see Hutto, 2007, and chapter 7 of this book) is relevant to agentive action involving higher cognition, but it does not tell the whole story. Much of the work done by Gallagher (e.g. 2005), Damasio (e.g. 1999) and others in related fields, however valuable it may be, also focuses on rather low-level sensorimotor processes, with limited relevance for higher cognition.

As a final mention, Thompson (2007) features a noteworthy attempt at correlating phenomenology and neuroscience from the enactive perspective (extrapolated from e.g. Varela et al (1991), Noë (2004)), and his project could, if developed further, have very interesting and significant consequences for my own approach, but for now it too lacks a sustained address of the problems of fitting higher cognition into the embodied and embedded cognition paradigm.

All the projects mentioned are important and valuable in that they foster awareness of the inadequacy of the old (i.e. cognitivistic) ways of conceptualising the mind. However, more than a few of these theories involving embodied and embedded cognition *shortchange the mental*. In their drive to define cognition in terms of fairly basic sensorimotor processes, combined with the abolition of as much internal representation as can be mustered, the inventors of these theories generate a characterization of the mind which does not appear to include a clear notion of what to do with higher cognition. In this book, I intend to offer a few suggestions on how we might be able to make some headway on this difficult terrain, with the explicit intent of presenting a model that is compatible with these promising lines of research on 'lower' forms of cognition.

In essence, I will provide an embodied and embedded theory of concepts - or, more precisely, of the concept 'concept'. This is useful because 'embodiment/embeddedness' and 'concepts' appear to be notions that are difficult to reconcile. That is, concepts are traditionally thought of as important building blocks of mental events or entities, and accounts along these lines usually reside squarely within the old paradigms. Thinking of concepts as components of thoughts feels quite natural if you define thoughts as complex symbolic representations, somehow occurring in the brain: a possible theory is then that concepts are more basic symbols that those complex symbolic states are composed of, and they can be combined in ways that adhere to specific language-like rules. Analytic functionalism, which does not endorse the symbol-based account of computationalism as sketched above, has well-developed ideas about concepts in terms of representations and representational structures as well. However, most embodied and embedded theories of cognition define mental terms in non-classical ways, for instance by involving extramental properties in their explanations of cognitive processes, often even by discarding the classical varieties of representations altogether.

The problem then is: what status should we award the concept 'concept' once we accept a theory which does away with these standard accounts of cognition and/or mental representation? So this is my central question:

[Central Question] How can we understand the concept 'concept' in an $E_{(i)}C$-appropriate fashion?

Hence, my intent is to explain the concept 'concept' in a way that fits with ideas about embodied and embedded cognition: this book's main focus is to provide a set of tools with which to specify what a 'concept' is, given the epistemological and ontological implications of theories of embodied and embedded cognition.

Towards that end, I will construct a model called the 'Radicality Manifold', which is intended to be a framework that describes concepts in an $E_{(i)}C$-appropriate fashion. The central idea will be to acknowledge the relevance of bodily influences as well as influences from the physical and social environment; collectively, these influences realize concept-involving behaviour. As such, I hope this book is a first, tiny step: an empirically informed conceptual analysis yielding constraints on a model of perceptuo-cognitive agent-world interaction, with the explicit intent to include perception, socio-cultural interaction and higher cognition.

The end result will be an $E_{(i)}C$-appropriate theory of what concepts are, and I believe the model will be flexible enough for the various flavours of $E_{(i)}C$ to adapt it towards their own ends. However, my own focus wil be mostly *enactivistic* ($E_{(A)}C$): the basis of the model (Thelen et al.'s (2001) dynamic movement planning field), Evan Thompson's theory of colour perception (which I use to expand Thelen et al.'s model) and the basic concept definition I utilise (which is ability-based) all fit in that general corner of $E_{(i)}C$. However, I will also incorporate a (very particular) notion of representation into the model, which is something most enactivists might be less enthused about.

The reason for this latter inclusion is the following: an important set of constraints, forming a kind of push-pull system of two opposing forces, concerns *phenomenology* and *representation*. These two notions do a lot of work in many classical theories, and are sometimes excorcised by embodied/embedded theorists. However, they appear very useful, and possibly essential, to accounts of higher cognition: phenomenology is about essential features of human experience, and representation figures heavily in theories about higher cognition. So, it seems a cautious handling of these two issues is in order; the issue is complicated because of the kind of use I wish to make of these notions. On the one hand, I want the model to be developed in this book to be phenomenally appropriate - that is, there needs to be a useful role to perform for phenomenal judgments - and this implies the necessity of at least some modicum of *internalism*. On the other hand, I am convinced that the warnings, voiced by followers of Gibson (1979), against excessive use of representations in theories about

cognition, hold at least some water - and this inclination exerts an *anti-internalist* force. Representation will receive attention in chapter 7; some modest clues about the role of phenomenology will be given in sections 6.6, 6.9 and 10.4.

1.4 - The Book's Structure

The main steps I will take throughout this book in order to reach the specified goal (an $E_{(i)}C$-appropriate theory of concepts) are the following. I will start with discussing the main theories about concepts, and establishing that they are insufficient as theories of concepts for various reasons (in sections 2.2-2.4). Following this discussion, I will introduce a successful *pro*ponent of the $E_{(i)}C$-approach, more specifically the $E_{(A)}C$-approach, which involves a model by Thelen et al. (2001) already mentioned above: it applies dynamical systems theory to cognition. However, a shortcoming of this model is that it merely describes (rather than explains) cognition-involving behaviour.

To make the step from behaviour and basic sensorimotor situatedness to concepts, we need clues about how the two hang together. To do this, I utilise the phenomenon 'colour' as a case study. Why pick 'colour'? And why spend so much time on it? I have two reasons for this.

I picked the colour case because in it, many of the most important themes of philosophy of mind are represented: given the many different processes involved in generating colour vision, the study of colour can be said to contain a *microverse* of the philosophy of cognition. Various issues in the philosophy of colour, and the various positions available within it, concern agentive action and interaction, mental states and phenomenology, social influences and microphysical processes, computational representation and embodied and embedded perception and action. That is, both in terms of the theories that are invoked to explain aspects of it, as well as the kinds of processes and properties that are involved in actually perceiving colour, the phenomenon of colour is a peculiar kind of in-between, frustratingly slippery in a conceptual sense, but exceedingly intriguing and useful as a test-case in the philosophy of cognition. Furthermore, colour vision is almost never merely perception, but almost always also involves decisions, and behaviour based on such decisions (involving whether fruit of a particular colour is safe to eat, or whether an object of a particular colour and visual texture is a dangerous predator or not, or whether a light of a particular colour means 'walk across the street' or 'wait for other traffic to pass', and so on). Because the sensorimotor contingencies of colour vision are fairly well researched, and because the connection to these higher-order processes is so natural, the case of colour vision is a good one to try and use in my own project.

The second reason, and this is perhaps the most important one, concerns my conviction that some of the solutions to persistent problems that have been constructed within the colour case, are applicable to the broader issue

of $E_{(i)}C$, or allow for a smooth and coherent extrapolation. I will discuss two complex controversies from the philosophy of colour vision, as I believe the solutions I will arrive at regarding the problems in question will provide me with the tools I need to construct an $E_{(i)}C$-appropriate theory of concepts. More specifically: in chapter 4, I will first address the controversy surrounding basic colour terms: is the structure of colour words in the language one speaks relevant to the way in which those colours are actually perceived? Relativists say yes, universalists say no. My suggestion will be that an intermediate position regarding the linguistics of colour is the way to go, accentuating the claim that embodied and embedded perception and cognition take place in a convergence zone of many different influences and forces, and no black-or-white solutions are available. This discussion will address the way in which an agent is embodied and embedded in his *socio-cultural* environment. The main problem to be solved then will be to provide the proper context (situatedness) for colour cognition. Theories of ecological colour provide exactly that ecological niche-based information. That is, next I will compare and contrast Evan Thompson's $E_{(A)}C$-compatible, *ecological* theory on colour perception with Roger Shepard's computationally inclined account; my conclusion will be that a true relational (rather than either subjectivist or objectivist) and $E_{(i)}C$-appropriate theory will need to take cues from *both* these theories. This particular discussion will address the way in which an agent is embodied and embedded in his *physical* environment. Hence *together*, the ecological and linguistic discussions describe the full range of embodiment and (physical and social) embeddedness of agents, at least inasmuch as colour is involved.

As a very substantial and significant bonus, the account of colour categorization that I will distill from the universalism-versus-relativism debate (in section 4.3), forms the backbone of my theory on *concepts* (chapter 6 and on). First, the case of colour perception will be used to illustrate the notion of a 'complex concept'; second, I will discuss 'Superposition Theory of Complex Concepts' (SToCC), a theory that strives to provide an appropriate characterisation of this type of concept, and resembles aspects of (but crucially differs from) Prototype and Theory theories regarding concepts; third, I wish to argue in favour of the idea that many more important concepts are complex in this way; and finally, I will introduce the *Radicality Manifold*, a model which generalises SToCC for concepts in general.

This theory of concepts, combined with the work on *describing* behaviour from Thelen et al.'s use of dynamical systems theory (section 3.2), will be refined and expanded to include various forms of embodied and embedded *concepts* (chapter 6 and on). This process of refinement will include a discussion about how representation - an important component of cognitivistic theories - might be used in $E_{(i)}C$-appropriate theories.

The evaluation of SToCC and the Radicality Manifold framework will occur in chapter 10, by comparing it to a series of existing theories of concepts, most notably the 'conceptual spaces'-theory by Peter Gärdenfors (2000)

and 'proxytype theory' by Jesse Prinz (2002). This final chapter will also contain a concrete application of my theory of embodied and embedded concepts, when I use it to analyse concept-based early childhood education.

A project like this could fill a library, or at the very least a rather portly book, so obviously it will not be possible to afford all the problems and theories the space and attention they require in this not-so-portly book. Above all, the ideas to be presented in what follows are intended to spark other ideas, and they should be read as a cursory overview of the humble beginnings of a theory about higher cognition in an embodied and embedded context.

In essence, the model to be developed is intended to generate an *epistemic* claim. It offers a suggestion on how to relate important data-domains (i.e. involving concepts, behaviour, biomechanical properties and environmental affordances) to eachother. This book is merely a sketch, a provisional and hypothetical framework resulting from philosophical analysis, that might have empirical consequences. The result will be a new model - a new metaphor, in a way - to talk about concepts within an $E_{(i)}C$ framework.

= =

[SUMMARY of chapter 1 AND PREVIEW]

The classical, 'Cartesian' way of thinking about cognition - of the mind being an entitity that has properties which set it apart from physical entities - is losing popularity amongst philosophers and psychologists. An important associated view is 'cognitivism', which involves the claim that cognition is to be defined in terms of *internal*, often representational processes.

An alternative view is 'embodied, embedded cognition', which suggests that a mental state is to be defined in relation to many relevant *extramental* properties, properties of the agent's body and affordances of his environment in particular. There are various flavours of embodied, embedded cognition:
*Embodied (cognition crucially involves bodily processes; notation: $E_{(B)}C$);
*Embedded (cognition involves interacting with an environment;
notation: $E_{(S)}C$);
*Enactive (cognition is an active, dynamic, behaviour-based process; notation: $E_{(A)}C$);
*Extended (processes in the environment form part of the cognitive process; notation: $E_{(X)}C$);
*Encultured (cognition depends on cultural processes for support; notation: $E_{(C)}C$).
General notation for 'embodied, embedded, etcetera'-theories: $E_{(i)}C$

An important component of many theories of cognition is a theory of *concepts*. This book is intended to provide a way of thinking about concepts that fits in with theories of embodied, embedded, enactive, extended and/or

encultured cognition. So this is this book's central question: **How can we understand the concept 'concept' in an $E_{(i)}$C-appropriate fashion?** This is an interesting project because 'concepts' as building blocks of thoughts are relatively easy to integrate into a cognitivistic story about cognition - for instance, if thoughts are symbolic structures, concepts can be more basic iterations of such symbols -, but it is much less obvious how to think of concepts within $E_{(i)}$C theories. The main focus of the theory to be developed in this book is enactive ($E_{(A)}$C). One of the main problems then becomes that many $E_{(i)}$C theories are quite successful in explaining basic sensorimotor agent-environment interaction, but are less adept at accounting for actual thinking, which is where many cognitivistic theories and their ideas about concepts are most effective.

Chapter 2 will discuss several standard theories about concepts, and their problems. Chapter 3 will introduce a useful, $E_{(A)}$C theory to describe fairly basic cognition-involving behaviour: Thelen, Schöner, Scheier and Smith's (2001) dynamical movement planning field. Chapter 4 is about colour perception, because in that field of study there are models available which suggest a connection between an agent's basic sensorimotor contingencies (the way in which the properties of his retina and the subsequent neural processing help him make distinctions between colours) and more advanced colour-related behaviour (for instance in terms of the cultural significance of certain hues, all the way up to scientific concepts about what colour itself is). The goal of this book is to adapt this connection - i.e. between basic sensorimotor properties which inform dynamic behavioural profiles on the one hand, and more advanced colour-related behaviour on the other - into an $E_{(i)}$C-apropriate theory of concepts in general.

= =

[2 - Introduction to concepts]

2.1 - General Properties of Concepts

The first step towards a theory of concepts is to provide a general definition that tells us what a 'concept' is. I will have the explicit goal of formulating this definition in terminology compatible with theories of embodied and embedded cognition. The first half of this first step is already quite a task, for there is little agreement about what a concept is, what it should be able to do, or what kinds of explanations are to be possible because of it. Georges Rey (1994) advances the thesis that there have been a number of quasi-successful attempts at determining what kind of a thing a concept is, but none of them have been able to offer a coherent response to any and all scrutiny. The concluding paragraph of his text offers a clear overview of the problem-space, and it deserves to be quoted in full:

"We might summarise the present situation with regard to candidates for 'concepts' that have been discussed here as follows: there is the *token representation* in the mind or brain of an agent, *types* of which are shared by different agents. These representations could be *words*, *images*, *definitions*, or *'prototypes'* that play specific *inferential roles* in an agent's cognitive system and stand in certain *causal* and *covariant relations* to phenomena in the world. By virtue of these facts, such representations become associated with an *extension* in this world, possibly an *intension* that determines an extension in all possible worlds, and possibly a *property* that all objects in all such extensions have in common. Which of these (italicized) entities one selects to be concepts depends on the explanatory work one wants concepts to perform. Unfortunately, there is as yet little agreement on precisely what that work might be." (Rey 1994, pg. 192)

As this quote demonstrates, there is little agreement about what concepts are in an ontological sense. It is, however, possible to make some suggestions about what kinds of abilities are associated with having concepts. It is possible to claim that such abilities contain three important aspects: (1) recall of the past, (2) conceptual categorization and (3) inductive generalization.

(1) - *Recall of the past*: An important aspect of a conceptual ability involves being able to act and react to an object or situation in a way that transcends the non-reflective immediacy of contextualised interaction. In other words: having a concept means being able to remember certain key facts or regularities about whatever the concept is of, which means that in such a case one tends to act and react differently than creatures without (or with significantly fewer) mnemonic capacities, who respond in ways that lie closer to the push-pull character of direct sensorimotor interaction.

However, it *is* important to make a distinction here, to pull apart (at least conceptually) the notions conceptual *ability* and the concept-involving ability *recall* of the past. The ability to remember (information about) the past is a

different kind of ability than knowing how to perform a particular (bodily) task. The former involves declarative memory, whereas the latter involves procedural memory. Declarative knowledge, or knowing-*that*, requires a certain level of conceptualization, being to some extent an abstraction that transcends the specific context in which it was learnt. Procedural knowledge is much more context-bound, being a partly or wholly intuitive, non-conscious knowing-*how*, that cannot (and usually *need* not) be made explicit except by simply performing the action in question. I want to claim that declarative memory (as the recall of declarative knowledge) can only occur as an extension of a substrate of embodied procedural recall-abilities, but that does mean that the former is an ability over and above a mere tendency towards certain patterns of sensorimotor agent-environment interaction: it involves the structured and in some sense decisive re-occurrence of such patterns.

Mandler (2007) notes that children as young as nine months are capable of remembering reliably how to work a particular toy after only having been *shown* what to do 24 hours prior; even some six-month-olds could remember what they were shown the day before. Carver and Bauer (1999) report many of the ten-month-old children they tested were capable of remembering a novel sequence of two actions a full month after witnessing a demonstration. Mandler's (2007) conclusion is that mnemonic abilities - hence, a primitive form of that particular component of 'adult' conceptual abilities - appear around six months of age, and have usually developed into a workable skill around ten months.

(2) - *Conceptual categorization*: An aspect of conceptual ability is the classification of objects in taxonomic structures. Understanding that cats, dogs, eagles and ants are all animals, and that animals are to be differentiated from (say) machines, hence having a grasp of the relations in which specific objects stand in different organizational structures, will influence the behaviour one unfolds in relation to those objects. In experiments carried out by Mandler (2007), even seven-month-old children were capable of differentiating between cars and planes. Contrary to popular belief, basic-level concepts (e.g. tables and chairs) were not the first ones to be formed - rather, superordinate categories (e.g. furniture) emerged first, with even children of eleven months failing to show further differentiation for this particular example (i.e. furniture), and only distinguishing dogs from cats in the animal paradigm; more detailed distinction abilities had not yet formed.

(3) - *Inductive generalization*: another important component or aspect of conceptual ability involves being able to use particular sets of criteria, causing objects to be grouped together or distinguished in ways that diverge from standard categorization based on visually detectable features. That is, having a concept of something means, at least in part, being able to act or react in relation to that object based on knowledge about non-obvious features or context-dependent interpretation of such features. Such features include, for instance, an object's function (forks and dinner plates look

rather different but are both meant to assist in eating), or its origin (the ceiling painting of the Sistine Chapel and the Piéta in St. Peter's Basilica were both made by Michelangelo), habitual location (bars of soap and razors are both usually found in the bathroom), biographical relevance (i.e. based on idiosyncratic associations derived from some past experience, for instance the first encounter with said object, or the way in which the agent learnt about the concept) or some other property that somehow pertains to what people might call the hidden 'essence' of the object (caterpillars and butterflies belong to the same biological cycle).

Over the centuries, there have been several theories about what kinds of things concepts are, given the tasks they were supposed to perform, such as the ones above. The purpose of this book is to provide an account of concepts that is compatible with $E_{(i)}C$, and that is capable of dealing with problems that plague other theories of concepts. Therefore I will first discuss three standard theories of conceptual structure: the classical theory, prototype theory and theory theory.

2.2 - The Classical Theory of Concepts

It is possible to summarise the central tenet of the Classical Theory (CT) of concepts as follows:

"Most concepts (esp. lexical concepts) are structured mental representations that encode a set of necessary and sufficient conditions for their application, if possible, in sensory or perceptual terms." (Laurence and Margolis, 1999)

In its strong, empiricist form, CT suggests it should be possible to reduce a complex concept to a collection of simpler concepts, eventually decomposing all the way down to the kinds of purely sensory terms that would occur in the 'observation sentences' or 'protocol sentences' the Logical Positivists wished to found knowledge on. An example of a somewhat weaker form of CT states that some irreducible, primitive concept-components might represent functional, social or cultural features - the demand that every concept ultimately reduces to nothing but sensory features is dropped.

The categorization of objects as falling under a specific concept in CT occurs by tallying the observable features of the object, and cross-checking it with the list of necessary and sufficient features associated with a given concept: for instance, an object is categorised as a rabbit just in case it is a smallish, furry animal that hops along and has elongated ears, a bushy tail, pronounced front teeth, etcetera.

A more careful look reveals several problems for CT (Laurence and Margolis 1999); I will highlight four of them. The first problem, and a rather serious one at that, is that for the vast majority of concepts, it is extremely difficult and perhaps even impossible to provide definitions in terms of an

exhaustive list of properties and features some concept is supposed to contain. In other words: concepts, in the majority of cases, do not have the kind of definitional structure CT needs them to have.

Another problem CT must face involves the possibilitie for dissociation between concept possession and the correctness of concept use. CT is a descriptivist theory about concepts, stating that to know a concept is to know its meaning and referents. The main argument against this kind of account is the idea that one can be wrong about what kinds of things a specific concept picks out, even though one can still possess that concept. For instance, superstitious beliefs about the origins of some illness do not negate the possibility of actually having the concept of that illness: the very possibility of saying those superstitious beliefs fail to pick out the *real* (i.e. modern, scientifically validated) cause of the illness presupposes the notion that, despite these different explanatory accounts, the antiquated and modern concepts of the illness are one and the same.

CT has difficulties accounting for the fact that a fair number of concepts are decidedly fuzzy. On CT, it should, in principle, be possible to determine category membership for some exemplar unambiguously, because each concept is supposed to cover a sufficient and necessary set of defining features. However, many objects disallow such a problem-free classification: for instance, Laurence and Margolis (1999) note it is not immediately obvious whether a carpet is a piece of furniture or not.

The final counter against CT I shall mention concerns socalled 'typicality effects'. CT does not have the resources to explain why some examples would be judged more typical of a given category than other examples – this critique is due to the proponents of Prototype Theory (to be discussed next), where the explanatory potential regarding typicality effects is an important selling point.

2.3 - Prototype Theory of concepts

The Prototype Theory of concepts (PT) was developed by Eleanor Rosch and her associates (e.g Rosch 1978), and was intended to provide answers to the problems that haunted CT. From the many different flavours of PT, the following core idea emerges:

"(...) most concepts - including most lexical concepts - are complex representations whose structure encodes a statistical analysis of the properties their members tend to have." (Laurence and Margolis (1999), pg. 27)

This notion allows for the fact that some things may be better examples of a particular category than others. This phenomenon can be explained by the idea that some concept's ascription to a given object is not a matter of that object possessing all the necessary features included in the concept (as per

CT), but merely possessing a sufficient number of them, and sufficiently many important ones. As Laurence and Margolis put it:

"(...) if BIRD is composed of such features as FLIES, SINGS, NESTS IN TREES, LAYS EGGS and so on, then on the Prototype theory, robins are in the extension of BIRD because they tend to have all of the corresponding properties: robins fly, they lay eggs, etc. However, BIRD also applies to ostriches because even though ostriches don't have all of these properties, they have enough of them." (Laurence and Margolis (1999), pg. 27, 28)

PT improves upon CT by being able to deal with some of the counter-arguments we described earlier. For instance, PT circumvents the first problem for CT mentioned above (that it is impossible to provide an exhaustive list of features) by acknowledging that concepts do not have the 'classical' definitional structure.

The main advantage offered by Prototype Theory is its elegant account of categorization, which in its general form is cast in terms of similarity judgments of a category representation and an exemplar representation. If the balance of weighted feature matches and differences exceeds a particular threshold value, membership of the exemplar is assured. The most typical members of a particular category will obtain the highest relative scores - hence, Prototype Theory naturally accomodates typicality effects (i.e. a robin is usually considered to be a better example of the concept 'bird' than a penguin). Prototype theory uses fuzzy set theory to characterise these typicality effects, and the behaviour of categories thus structured under conceptual combination[NOTE 7].

Despite its considerable advantages, PT also needs to face a number of problems. For instance, Armstrong, Gleitman and Gleitman (1983) suggest there is a methodological problem in the way PT suggests we make the step from typicality effects for concepts to a more general theory about prototype structure. Even though something either is or is not an even number, and all of the test subjects consulted by Armstrong et al. agreed that the well-defined category EVEN NUMBERS is not graded, still there proved to be a tendency amongst them to consider 8 to be a better example of an even number than 34. This would mean that PT would have to explain how typicality judgments, apparently, are not (or not *always*) about degrees of membership (after all, technically there should not be a difference in membership status between 8 and 34, yet the test subjects state otherwise). However, it is exactly this linkage (between typicality effects and prototype structure) that the supporters of PT wish to employ to establish assertions about the prototypical structure of the concept in question.

Another problem for PT is that the kinds of features that can come to be encoded in prototype representations of some concept do not necessarily pick out the concept's correct extension. Armstrong et al. (1983) note that a three-legged, tame, toothless, albino tiger is still a tiger (despite not possessing many of the properties thought by most people to be

characteristic of tigers), whereas a very convincing toy tiger (which might possess many of the appearance-based properties associated with the concept 'tiger') is not actually a tiger. As with CT, the problem is that possession and application of a concept can be out of synch with what the theory (PT in this case) says the concept is supposed to be.

There are two additional, and more serious problems that can arise for PT: (1) for a large number of concepts, test subjects fail to isolate any typicalities (the so-called 'missing prototypes'-problem), and (2) attempts to combine graded extensions run into serious problems. Both these issues will be discussed more extensively in section 6.11.2, as I attempt to explain how my own theory, despite its similarities to PT, is able to circumvent these problems.

2.4 - Theory Theory of Concepts

The central notion of the Theory Theory of concepts (TT) is as follows:

"Concepts are representations whose structure consists in their relations to other concepts as specified by a mental theory" (Laurence and Margolis 1999, pg. 47)

Hence the criterion for subsumption of some object under a specific concept is not due to some intrinsic list of concept features, which is the case for the CT and PT, but rather is determined by the mental theory that someone might employ to explain what the concept is. As with the theory theory of mind, this 'theory' need not be an explicit scientific account, but can include a variety of knowledge-informed, more or less implicit categorization tendencies.

One of the advantages of TT cited by Laurence and Margolis is that this idea accomodates widespread views about psychological essentialism, i.e. that often judgments of category membership are not really about an object possessing a sufficient number of relevant properties, but rather that the object has a specific underlying nature or essence, which might be hidden. Consequently, someone who possesses a concept need not be aware of all the details of the underlying theory, but rather be disposed to the use of a so-called 'essence placeholder'. The main advantage of this strategy is that it makes epistemological sense: it comprises the intuition that objects that look similar probably share other, hidden features, and relatively often such an assumption turns out to be correct.

There are two main lines of criticism that can be directed against TT. The first underscores the problem that the aforementioned essence placeholders are too sketchy, that the lack of represented information (most test-subjects fail to give a clear account of what this essence consists of) disallows the concept in question to pick out an extension in a proper way. A related problematic issue for TT consists in the assertion that people might represent incorrect information as part of their essence placeholder.

In this case, too, the concept as it is held by someone might fail to pick out the correct extension. The second argument against TT involves the question how a concept can stay invariant despite changes in belief. In other words: how can people with different theories still be said to possess the same concept? Laurence and Margolis (1999) claim that, as yet, few coherent answers are forthcoming.

2.5 - Desiderata on a Theory of Concepts

The discussion in the sections above shows that the most prominent kinds of concept theories fail to account for certain key features of concepts. Obviously, we need a new theory, but which criteria should such a theory meet? The easy answer is that it should be able to deal with the problems, described above, that plague the existing theories. Explicit attention to the problems of the standard concept theories mentioned above, and the ways in which I feel my own theory can solve such problems, will be given in section 6.11. In a more general sense, Jesse Prinz suggests a good list of workable criteria at the beginning of his book on concepts (2002):

Scope: a successful theory of concepts should be able to explain a rather diverse range of concepts: phenomenal feels, formal notions, natural kinds, and so on.

Intentional content: concepts stand in for or refer to entities, processes and/or states-of-affairs in the world: the concept 'dog' is about actual dogs

Cognitive content: Prinz suggests that concepts need to have something like a Fregean *Sinn* (sense) apart from their *Bedeutung* (reference). That is, apart from their intentional content (the ability to refer to external entities), there needs to be some additional content-type - cognitive content - based upon which coreferential but psychologically distinct concepts can be individuated. Along these lines, Prinz suggests that concepts should be individuated through reference to both external entities and other concepts.

Acquisition: a theory will need to be able to account for the acquisition of concepts, in an empirically adequate manner. That is, a philosophical theory of concepts should be compatible with at least some account of the evolution of concepts and concept learning.

Categorization: a specific concept usually refers to objects that belong to a specific category. A proper theory of concepts should explain how such categories are formed, and how objects are recognised as belonging in one group and not another.

Compositionality: concepts can be combined to form more complex concepts - a theory should be able to account for this generativity of concepts. The content of the individual concepts and the content (both intentional and cognitive) of the more complex concept-components should

be related according to a compact suite of combination rules. This feature will help to explain the systematicity of thought.

Publicity: concepts can be shared by different agents, and can be used by one agent at different times. For different people to understand eachother, both intentional content and cognitive content should be sharable.

At the end of this book, in section 10.5, I will evaluate my own theory of concepts (to be developed in the pages to come) by determining to what extent it either meets these criteria, or makes a different tenable suggestion. Other tools I will use to test my theory of concepts include answers to the criticism on standard theories, as described above (in section 6.11), comparisons to somewhat similar theories (sections 10.2 and 10.3) and a description of a concrete application of my theory (section 10.4).

= =

[SUMMARY of chapter 2 AND PREVIEW]

It is highly difficult to provide a good definition of what a concept actually is: candidates are, amongst others, mental representations and abilities. Basic components of conceptual abilities are recall of the past, conceptual categorization and inductive generalization

The standard accounts of concepts discussed in this chapter are:
-Classical theory (a concept is a representation which encodes necessary and sufficient properties of an object);
-Prototype theory (a concept is a representation which encodes properties of objects in a graded fashion);
-Theory theory (conceptual structure is determined by a mental theory).
All these theories were explained to be insufficient for several reasons.

A proper theory of concepts should allow concepts to have appropriate *scope*, it should be able to explain how concepts can have *intentional content* and *cognitive content*, how concept *acquisition* occurs, how it can be that concepts allow for *categorization*, how simple concepts can be combined into more complex concepts (*compositionality*), and how it is possible for different people to hold the same concepts (*publicity*).

In chapter 3, the first steps will be taken towards an $E_{(\ i\ )}C$-appropriate theory of concepts, using a dynamical movement planning field as a description of basic sensorimotor behaviour.

= =

[3 - Embodied Dynamics]

3.1 - The First Step: Enactive Cognition?

In order to formulate a coherent $E_{(i)}C$-appropriate theory of concepts, I will first take a look at an existing $E_{(A)}C$-approach to cognition - cognition-infused *behaviour*, to be exact.

An important body of criticism on cognitivist/computational theories of cognition is to be found in the enactivist tradition; Varela et al. (1991) is a formative publication of this theoretical perspective, as is O'Regan and Noë (2001). Thompson (1995a,b) develops an enactivist theory of colour perception, and this account will play a large role in the pages to come. That is, later (in section 4.5), I will attempt to demonstrate that some ideas encased in the ecological theory of colour perception developed by Roger Shepard, offer interesting additions to Evan Thompson's $E_{(A)}C$-approach to colour perception. The claim will be that Shepard's specific approach, which happens to be computational, highlights the need to take into account certain structural aspects of the agent-environment interaction-dynamic. More specifically, his model contains a description of the interplay between the properties of an agent's perceptual system (including the properties of the phenomenal structuredness generated in perception) and certain structural invariants of the optic array. Regardless of the computational details of the model or the ontological status awarded to its components, the explicit linkage of what I want to call *affordances of environment and agent* comprises an important lesson for the $E_{(A)}C$-inclined to take into account. And much later, in section 7.2, I will visit Hutto (2006, 2007) for a critique that states that many enactivist theories *are not enactivistic enough*. I will not agree with everything Hutto says, but I will use his criticism to sharpen my own theory about $E_{(i)}C$.

3.2 - Modeling the Dynamics of Behaviour

As a first suggestion on how to describe behaviour, I will devote a fair amount of space to discussing the *dynamic movement planning field* model by Thelen et al. (2001)[NOTE 8], some aspects of which will help in the construction of my own theory (the Radicality Manifold, which is intended as an *explanatory template* for $E_{(i)}C$-related phenomena). The idea is that Thelen et al.'s model provides a way to describe behaviour in an $E_{(i)}C$-congruent fashion, but that its explanatory power is curtailed because it, as it stands, merely applies to a fairly basic cognitive phenomenon, and because it goes too far in downplaying the role of psychology (i.e. of accounts actually involving 'the mind') in providing valid explanations. Part of the work towards constructing the Radicality Manifold-model involves adapting Thelen et al.'s model in a way that corrects these shortcomings.

Thelen et al.'s (2001) model is a fairly successful application of the tools of dynamical systems theory[NOTE 9] (DST) to a problem involving a cognitive process (DST-C) - in this case, the ‘A-not-B error’. The ‘A-not-B error’ was

observed and described by Piaget, and involves a fairly common behavioural quirk of children between the ages of 7 and 12 months, who, after successfully retrieving a hidden toy from location A, persist in reaching towards that location even after witnessing the toy being hidden at location B. Traditional explanations are cast in terms of poorly developed object representations in the children making this error, but the model Thelen et al. propose does away with internal representation altogether, and is intended to yield novel predictions. In this model, tasks like perceiving, planning and remembering, typically described in predominantly mental terms, are effectively absorbed into the mathematical tools used to describe the movements[NOTE 10]. This method of description underlines the idea that cognition and bodily action are inextricably linked: cognition is *embodied*, and is to be understood in dynamical terms[NOTE 11].

As an explicitly stated goal, Thelen and colleagues strive to abolish theories about cognition and action that involve *internal representation*:

"Our message is: if we can understand this particular infant task and its myriad contextual variations in terms of coupled dynamic processes, then the same kind of analysis can be applied to any task at any age. If we can show that "knowing" cannot be separated from perceiving, acting, and remembering, then these processes are always linked. There is no time and no task when such dynamics cease and some other mode of processing kicks in. Body and world remain ceaselessly melded together." (Thelen et al., 2001, p. 2)

Thelen at al. appear to suggest that if their model is successful in describing the child's behaviour and in yielding new and improved predictions and explanations, this would provide a kind of proof by way of demonstration that many theses involved in $E_{(a)}C$ are coherent, i.e. that many (or possibly all) cognitive processes are to be understood in terms of embodied, sensorimotor interaction dynamics with the environment (enaction). They forge an explicit commitment to the thesis, present in most forms of $E_{(A)}C$, that internal representation is much less important in the explanation of cognition than cognitivist theories suppose - in fact, they say we can do without reference to internal representation.

It deserves to be noted that their claim that embodied dynamics is always involved in cognition, i.e. that 'there is no time and task when such dynamics cease', does not necessitate the conclusion they wish to support, namely that no other mode of processing could ever play a role. Part of my suggestion will be that representation (and sometimes even *symbolic* representation!), does have a role to play, but that these mental phenomena should be understood as *limit cases of a spectrum* that *also* comprises the basic $E_{(i)}C$ dynamics they concern themselves with. In brief, I *agree* with the fact they afford embodied dynamics a central spot in explanations of cognition, but I *contest* the suggestion that their approach would be capable of telling the *whole* story.

The development of this suggestion will have to wait, because first I will take a brief look at the model itself[NOTE 12]. The model contains several essential elements[NOTE 13]:

(1) the relative ambiguity of the task input: the lids of the hiding locations look identical;

(2) the relative strength of specific input: a brightly coloured toy or a cookie is more interesting to a child than a blandly coloured toy;

(3) there is an enforced delay between the task input and the onset of the child's action, but obviously the dynamics of the processes contributing to the child's actions are not paused during that time: things continue to happen;

(4) the A-not-B error arises in the act of reaching;

(5) the motor memory of the initial (successful) reach towards location A influences the subsequent tasks;

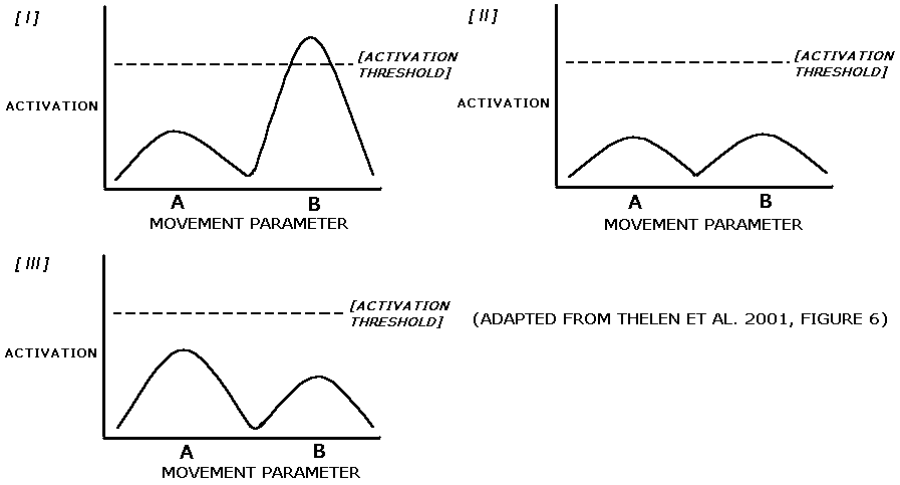
(6) children grow out of making the A-not-B error: one hypothesis is that changes in one or more of the relevant parameters contribute to this development.

Going against the grain of prior explanations, Thelen et al. hypothesize that the A-not-B-error *does not* arise due to an inability to access or utilise the proper internal representation mechanisms, but rather should be analysed in terms of the outcome of 'a motor planning process that is part of a dynamic perception-action loop' (2001, p. 11). As a generalisation of the six elements described above, four theses are claimed to hold, and these form the foundation of the model of the A-not-B-error they construct. Quoting:

"(1) actions are planned in movement parameter space; (2) the plans are continuous and graded in nature; (3) plans evolve under continuous perceptual influence of both task and cue; (4) the system has history." (Thelen et al., 2001)

At the core of the actual model lies a continuous movement planning field (see figure 2): a mathematical model which simulates the behaviour of the child. One of the two regions in the field spiking beyond a certain threshold value specifies the child's reaching for either A or B. A super-threshold spike, specifying a reach, must build up over a period of time.

Figure 1 shows three different activation scenarios of this movement planning field: [I] shows how the activation exceeds the threshold value for a reach towards location B, [II] shows low activation levels across the entire field, hence there is no reach, and [III] does show heightened activation in some region of the field, but it is sub-threshold, hence insufficient to effect a reach.



[Figure 1: movement planning field used by Thelen et al. (2001)]

Previous states of the system influence the current dynamics: the input to the system has three components, and concerns not only information about (1) the task structure and (2) specific input nudging the system towards either A or B, but also (3) the memory of earlier attempts to reach for the toy. These three types of input are sources of bias to the field. The *task input* specifies the layout of the task space (for instance, the location of the two hiding wells - say, two compartments in a box - , the colour and dimensions of that box, and so on), which is usually invariant throughout a single trial. The *specific input* concerns the attention-demanding acts of the experimenter; that is, the measure in which attention is drawn towards the toy (for instance by waving it around, or tapping it on the edge of the box with the hiding wells). The *memory input* is crucial, for knowledge about previous reaches influences the system's performance, not only affecting the probability of reaching towards A or B after the experimenter's cue, but also affecting spontaneous reaches. The model's performance here is congruent with experimental observations of children exhibiting the A-not-B error.

A central feature is that the field exhibits cooperativity: to help generate a single response with complex input, regions of the field that lie close together are mutually stimulatory, while inhibiting more distant activation. In simulating the A-not-B-error with the model, parameter values were chosen to help the model reflect the real system performance (i.e. the actual behaviour of a child attempting to perform the task). Many of these values were held constant throughout simulations (to reflect, for instance, the invariance of the task field). Three factors were varied, and thus influenced the model's behaviour: (1) the specific input (either A or B); (2) a parameter expressing asymmetry in the task arrangement (whether one target, A or B, was significantly more interesting than the other, for instance a dull-coloured toy vs. a brightly coloured one); and (3) the cooperativity level of

the movement planning field (with a choice of two values, to express either a cooperative or non-cooperative regime). The resulting output of the model specifies a reach towards either A or B; a model of arm motor control is not included, and Thelen et al. suspect the specifics of such a model used could have a nontrivial influence on the behaviour of the system as a whole, but as an approximation, they assume the current model as it stands is sufficient to help them prove their point (contra 'cognitivism', and the use it makes of internal representations).

The model incorporates a few novel elements. The cooperativity of the field, for instance, is a crucial parameter, and apparently one that is expected to do a lot of the work in yielding the appropriately realistic simulation of the A-not-B-error. In several simulation trials, the difference between non-cooperativity and cooperativity of the movement planning field meant the difference between a notable lack of self-sustaining activation, and the emergence of strong, threshold-exceeding peaks, respectively. The difference between these two parameter settings is supposed to explain why younger children make the A-not-B-error, whereas older children do not: as the children grow older and more experienced in interacting with the world, the cooperativity of their movement planning field (which can be used to model their behaviour) improves. Another novel aspect of the model is the role of memory: rather than a store of knowledge to be accessed in a comparison of representations (of the memory and current input), or whatever other representational story one would wish to devise, memory is implicit in the dynamics of the movement planning field.

3.3 - Dynamics Deconstructed

I am quite sympathetic to Thelen et al.'s endeavour, and I agree that their model is successful, but only up to a certain point. Thelen et al. state that a simulation of various (measured) experimental results with their model does indeed demonstrate that the kind of dynamics captured by the model resembles the dynamics of the real system in several significant ways. One pertinent question then is: what is the status of this 'movement planning field'? And: what kind of predictions is this model supposed to generate, and how robust and specific can these be expected to be? One source of initial scepsis can be that the precision requirements do not appear to be very high: the model's output is a sequence of A or B choices, which only needs a general similarity to the real system's dynamics (due to the large influence of noise and general probabilistic inclination of subjects in the task described) to be considered appropriate.

But it is possible to develop a somewhat more substantial critique of the model: Thelen et al. do some of that work themselves. These are the limitations they understand their model to have:

(1) it captures the real system's behaviour, but the way it, its components or its dynamics are supposed to map onto or be related to the biology and

neurophysiology of the child in any other than the most general of fashions is not specified in any substantial way;

(2) it is incomplete, for it is not linked to an appropriate model of the child's reaching dynamics, i.e. the particular shape of the arm- and hand-motion executed by the child. Because of the stress Thelen et al. place on the embodied character of cognition, such an additional model's properties can have a profound effect on the functioning of the coupled system;

(3) the A-not-B-task involves actions geared towards both location and object identity (a child does not reach to a location in itself, but to *an object* at a specific location). The model is a simplification, since its output only allows one parameter dimension, namely location (either A or B);

(4) it only involves static visual targets, and does not incorporate movement, or, for instance, auditory stimuli.

These are, indeed, weak spots in the model, or at least marks of incompleteness, and the authors are practicing good science by conceding to their existence themselves. The most pressing concern from my perspective (i.e. in the light of the theory that I wish to develop) is the strong intuition that weakness (1) can be expanded with another element: their model does not offer any room to do justice to the idea that a good *explanation* of cognitive phenomena, especially more complex ones, often requires reference to mental states (such as beliefs and desires), for instance to help provide the appropriate justification - i.e. *reasons* - for certain actions. Perhaps explanations of the low-level actions involved in the A-not-B-error are indeed better off without excessive reliance on psychological processes (understood in terms of internal, possibly representational processing), but for more advanced action sequences, I would suggest that blocking out any and all possibility of referring to the mind behind the motion equates throwing out the child with the bath water.

My own suggestion on how to handle this issue will follow later - introducing and developing this suggestion is the one of the main goals of this book, after all -, but first a closer look at the problems of Thelen et al.'s model is warranted. Some of these problems are addressed in the peer commentaries to Thelen et al.'s target article. These commentaries are, for the most part, positive about the potential of the model and the improvements over the earlier iterations (e.g. as described in Thelen and Smith 1994), but often critical of details.

Markman (2001), for instance, says that the relevance of this kind of modeling to other, more complex kinds of cognitive behaviour is yet to be demonstrated. Furthermore, and this is a much more serious problem, it is as yet unclear how this particular embodied, non-representational use of DST-C on its own (i.e. as a methodology and associated ontology severed from related and potentially helpful disciplines such as connectionism) could yield psychological explanations rather than merely descriptions of

behaviour - this comment amounts to the accusation that Thelen et al.'s model is good old behaviourism in a new form, also raised by Eliasmith (1995, 2003), and is echoed in various flavours by Pelphrey and Reznick (2001) and Sophian (2001). Appropriate though it is, it should be noted this particular counterargument was more forceful against Thelen and Smith's less sophisticated work of the mid-1990s (e.g. Thelen and Smith 1994, Thelen 1995), so this particular research programme does appear to contain the momentum which might allow people working in this paradigm to meet this criticism sometime in the future. However, a lot of work still remains to be done, and my own model to be developed later in this book can be seen as a small contribution towards that end.

An additional problem is that there is a hint of circularity in Thelen et al.'s dismissal of the importance of representation - a hypothesis that is a cornerstone of their brand of DST-C:

"The model accomplishes all this without invoking constructs of ''object representation,'' or other knowledge structures. Rather, the infants' behavior of ''knowing'' or ''not-knowing'' to go to the ''correct'' target is emergent from the complex and interacting processes of looking, reaching, and remembering integrated within a motor decision field." (Thelen et al. 2001)

I have no quarrel whatsoever with the thesis that 'knowing' and 'doing' are intimately connected: I am quite convinced that it is true. However, the authors might be overextending their model's potency by making internal representation a factor of practically no significance whatsoever. Via their *a priori* adherence to a particular notion of $E_{(i)}C$ (one that I would characterise as closely allied to $E_{(A)}C$), this assertion was an *assumption* - if it is supposed to be a 'conclusion' to be derived from the simulations involving the model, this should not come as a surprise. This is because by design, the model depicts only the behaviourist aspects of the real system, and says little about the cognitive underpinnings of the movement-dynamics, nor does it rule out object representation or somesuch variety of internal processing as a byproduct of the embodied dynamics, or even as a mechanism that exerts a modicum of control. Thelen et al. disqualify representations as components of their explanation *by stipulation*, and rely on the success of their model to scaffold the thesis that even if one were to invoke representations, they would do no relevant work.

However, if this is how the model is supposed to work, here's the rub: the model mainly presents an after-the-fact *description* of behaviour rather than an *explanation*, except perhaps in the most abstract of forms: the relevant factors and parameters are explicitly chosen to have the model resemble behaviour that has already been observed. The case for the causal impotence of representation would have to come from the *explanatory* success of the model, but that is where the results are somewhat meager. Or, to reiterate an important qualifier (the red thread of this commentary) to this statement: I believe a 'representation-light' approach to low-level activities such as being engaged in falling prey to the A-not-B-error is

probably the right tack to take, and the success that the model *does* exhibit is sufficient to cement this belief, but this strategy does not suffice for more complex cognition-involving behaviour.

Mark Latash (2001), in his peer commentary to Thelen et al.'s 2001 target article, makes a similar point. Latash notes that the strong anti-dualism of Thelen et al.'s model and underlying theory 'explains' away all forms of mental activity per se in favour of their embodied account. However, this renders their model unable to account for some data. Latash found that in experiments where subjects were asked to practice the unnatural task of mirror writing, subjects would feel as if their hand refused to cooperate, a situation analogous to the uncooperative field setting of the Thelen et al. model that was intended to simulate the behaviour of infants prone to making the A-not-B-error. Uncoupling bodily and visual feedback from the mental guidance of the task at hand would often help overcome the impasse situations in the mirror-writing tasks, suggesting that in complicated cases, for instance requiring novel decisions or crisis management, something else is necessary to complement Thelen et al.'s scheme defining knowing in terms of perceiving, moving and remembering. To keep Thelen et al.'s model from being nothing more than a straightforward stimulus-response-description, some reference needs to be made to internal processing - imagination, perhaps. Still, Thelen et al. claim that there is no internal representation of significance. Later, when I turn to the ideas of Andy Clark concerning 'representation-hungry problems', this topic will be explored further (see section 7.4).

As a final issue with Thelen et al.'s model I wish to mention, Eliasmith (1995, 2003) accuses DST-C of adhering to views closely allied with behaviourism, and this criticism still applies. In this case, the a priori adherence to a fairly orthodox version of $E_{(A)}C$ professed by the authors plays such a major role, that it prohibits the abstract nature of the mathematics of DST from being translated into a more concrete account of cognition. It should be noted that, despite leaving the distinct impression it ultimately falls short, the model constructed in Thelen et al. (2001) *does* present a significant improvement over, for instance, the work of Thelen and co-workers in the mid-1990s, and the continuing evolution of the work in this field holds promise for the future.

3.4 - Newton's Curse

I would like to claim that the moral we should take home from this discussion is that Thelen et al.'s intriguing way of attempting to provide explanations of cognition (or cognition-involving behaviour) in terms of embodied dynamics and enaction can get us reasonably far, but not nearly far enough. To revisit one of my earlier comments (from section 1.3): I think something more is needed to allow explanations of more complex cognitive processes - actual 'thinking', rather than fairly basic sensorimotor interaction of agent and environment, however complex those interactions can already be. I do believe that Thelen et al.'s model is rich enough to serve as a

template for a more comprehensive model - my own 'Radicality Manifold' (RM) is an attempt to cash in on that potential. The RM is intended as an *explanatory template* for $E_{(i)}C$ in general, offering an overview of the components which are supposed to co-constitute a proper explanation of cognition in that sense, and the kinds of relations that are realised in the interaction between those components. This is the model that I intend to develop throughout this book, and Thelen at al.'s model to account for the dynamics of cognition-involving behaviour is a good starting point.

But first I wish to clarify the direction that I feel we should take, if Thelen et al.'s DST-C is to be understood as a jumping-off point. One of the central tenets of $E_{(i)}C$ is that many properties of agents cannot be understood via microphysical reduction, nor can they be characterised properly if divorced from their context. As a result of these tenets, $E_{(i)}C$ (especially $E_{(A)}C$ and the related DST-C) presupposes an account of causality that is problematic, or at the very least divergent from the traditional scientific norm. Rather than a sequential ordering of neatly delineated cause-and-effect pairs, many $E_{(i)}C$-approaches suppose that we should understand systems, or at least *living* systems, in terms of a circular interaction dynamic of properties and processes at different spatial and temporal scales (see also section 9.3).

The contributions of Galileo and Newton to physics seriously depleted our ability to understand the organic whole of a living system, the kind of causal dynamics it is immersed in, and the methods by which that dynamics is to be described and explained. Despite Isaac Newton's monumental achievements, and his undeniable status as one of the greatest minds ever to grace humanity, I will be ever-so-slightly blasphemous and call this problem *Newton's Curse*. In broad terms, I understand this to mean the following: the problems associated with the tendency to extrapolate an ontology from an abstract, structure-based metaphysics.

This problem has much older roots - it might even be possible to trace it back to the success of the early 'sciences' of the ancient Greek scholars: mathematics, logic and philosophy. The success of the first two of these disciplines' focus on form and structure, rather than content, and the predilection of the third for a thoroughly rational (rather than perceptual) approach to reality stunted the proper emergence of what we currently understand to be empirical science. In all three cases, empirical data was considered to be of much lesser importance than introspective research methods. Despite modest empiricist beginnings with Aristotle, and a stronger empiricist turn from the seventeenth century onward (Bacon, Hobbes, Locke, Hume), this broadly rationalist inclination remained of utmost importance, up until today (at least in folk-physics).

Some of the earliest proponents of this view were probably Pythagoras, with his mathematics-based ontology of the universe, and Democritus, whose efforts were geared towards reconciling the philosophies of Heraclitus and Parmenides. Heraclitus claimed that reality is in constant flux, and any and all understanding we might hope to appeal to are merely temporary

coagulations of this ongoing transformational process: Πάντα ρει και ουδην μένει ('everything flows, nothing stands still'). Parmenides, rather, came to believe that the appearance of change merely masks the true, unchanging character of reality (άληθεια, 'truth'). The clever solution Democritus arrived at was to posit the existence of unchanging atoms in ever-changing configurations. Furthermore, he intended the processes by which these configurations occur to be wholly deterministic: one state, out of necessity, would lead to the next.

Consequently, Democritus claimed that empirical data only exists as secondary qualities: whatever we perceive to be hot, or sweet, or of a particular colour, is only so by convention. Our senses cannot provide us with an accurate portrayal of what the world is *really* like, hence whatever knowledge we gain from perceiving is knowledge of a lesser kind; the real structure of reality can only be probed via our faculty of reason.

Democritus' achievement is highly impressive, especially considering how well the major themes from his philosophy - reduction and the epiphenomenalism of macrophysics and sensory properties, plus the primacy of reason - align with the conceptualisation of reality as it emerged much later from the era of the scientific revolution, from the contributions of Galileo, Newton and Descartes, in particular.

Despite the similarities of the Newtonian view with this classical theory, a lot of nuance of ancient Greek philosophy was lost. Juarrero (1999) notes that of Aristotle's four causes (material, formal, efficient and final), efficient cause has become most prominent in what could be characterised as the Newtonian orthodoxy, and which is still a major component of folk-physical beliefs. That is, the only kind of causality that is utilised in naive physical theories is 'billiard-ball causality', the kind involving objects bumping into eachother and transferring force and impulse. This aligns with, or might be co-constituted by, the deep-seated conviction that everything we see around us is nothing more than a collection of microphysical entities - something like Democritus' *atoms*.

The Newtonian view underscores this conviction, conceptualising causal primacy in terms of a reduction of wholes to parts, where the wholes are causally impotent epiphenomena, i.e. merely aggregates of microphysical constituents. Note how this view is still common today, i.e. as exemplified by Jaegwon Kim's (e.g. 1998) theories on supervenience. Block's (2003) objection to this view - that it runs the risk of 'causal drainage', of macrophysical powers being mere epiphenomena of constellations of powers contributed by microphysical realisers - uncovers the problem quite nicely.

This reductionistic view has the unpleasant side-effect (unpleasant at least for $E_{(i)}C$-supporters) of demoting *relational* properties of macrophysical entities to the status of *secondary* qualities. I will show (in chapter 4), an

$E_{(A)}C$-based critique of standard theories about colour involves exactly the claim that colours are actually relational, while the orthodox, Lockean view brands them secondary qualities of objects.

Galileo fostered the hypothesis that, for many practical intents and purposes, there was little harm in ignoring contextual effects, such as atmospheric friction for falling objects. Where the influence of the environment could not be ignored, a serious problem emerged: modeling the properties of the environment of some particle would require modeling the parallel states of all other particles, and this simply could (and can) not be done.

Newton proposed a rather clever solution: he conceptualised the influence of the environment on a particle in terms of that particle's change of state, under influence of some force. Chemero and Turvey (2006) note that this is part of a vast *explanatorily* reductive manoeuvre, encapsulated in Newton's second law[NOTE 14] $F(x, \dot{x}) = m\ddot{x}$. This manoeuvre includes a reduction of the number of variables describing a particle's state from (potentially) infinity to *two* (position x at time t, and velocity $\dot{x}$, collectively specifying the particle's *phase*). The effects of the environment are defined to be proportional to the particle's acceleration $\ddot{x}$, and modulated by the particle's mass.

This clever solution did have a number of complicating consequences, amongst them being the proclivity to model causal processes of systems in terms of a sequence of state transitions, and positing as non-entailed those influences that are of an origin external to the modeled system.

The former point, combined with atomism, has the peculiar consequence of supporting an ontology in which time is, in principle, reversible[NOTE 15]. This suggests that the diachronic properties of systems as defended in (for instance) $E_{(A)}C$ require a non-Newtonian viewpoint - that is, explaining such properties requires a different conception of causality (see section 9.3). The latter consequence of Newton's clever solution as described above can be read as a negation of the $E_{(S)}C$-component of the general $E_{(i)}C$-inclination: situatedness (embeddedness) is not of explicit influence on what a particle/entity/agent is to be defined as being.

Descartes' viewpoint fits into this perspective quite readily by defining the world and its inhabitants in materialistic, atomistic terms, and then coming to the conclusion that this simply will not do when describing human beings: there should be something other than the body that is nonetheless essential to man... and that, according to Descartes, is the immaterial soul. The motion of the material body has material causes, but, barring awkward 'solutions' involving the pineal gland as the locus of matter-to-mind interaction, Descartes has to claim that the workings of the immaterial mind can only be self-caused, in virtue of it having free will (Juarrero 1999).

When one adopts the Newtonian account, including atomism and determinism, one runs into major problems when Cartesian substance

dualism is discarded, especially when this is done half-heartedly. That is, the success of the Newton-inspired sciences promulgated a view of reality with no clear place for the intuitions of many concerning the properties of the human mind (e.g. the ideas developed in phenomenology), and the problems of this view were compounded by the conviction that the mind *does* need to be accounted for in a manner congruent with those intuitions. One aspect of this odd confluence of Newtonianism and Cartesianism includes the idea of the mind as an entity, property or process that somehow plays by its own rules, and because of that idiosyncratic character requires *representations* as the internal effects of external causes, as *stand-ins* for those distal processes.

The formal methods of description of the Newtonian account can be recognised in a related approach to explaining the mind: computationalism. Mental processes are described in terms of functional properties and regularities - a particular mental state has the function of being an intermediary between an input state and an output state. Where Newton describes material processes in terms of sequences of a phase state of some particle (plus an environmental force) entailing a particular other state, the computational approach to the mind depicts functional state transitions as transformations of symbolic structures in accordance with certain rules into other such structures.

These Newtonian/Cartesian implications are examples of a style of conceptualising the mind that is most vehemently opposed by supporters of $E_{(i)}C$. In this light, it might count as odd that DST-C, which I classified as an endeavour closely related to $E_{(A)}C$, uses Newton's differential calculus as a base component of its models (see section 3.3). I believe that DST-C has the resources to defuse at least some aspects of Newton's curse, but certain other problems still remain.

First, the positive differences: DST-C promotes the claim that standard computationalist approaches consistently fail in accounting for the temporal dimension. DST-C is intended to bring diachronicity back to the forefront, by claiming that the dynamics of a dynamic system depend, to a large extent, upon the history of the system (Van Gelder and Port, 1995). Furthermore, DST-C (and with it virtually all variants of $E_{(i)}C$) is decidedly holistic, supporting, as noted above, an account of causality that involves parts and wholes, and the properties connected to those parts and wholes, interacting with eachother in complex ways.

However, I believe that for DST-C, Newton's Curse, broadly conceived, still remains. That is, the models of DST-C still exhibit the tendency to extrapolate a qualitative ontology from an abstractly quantitative metaphysics. Using such models as descriptions and tools for prediction is a valuable excercise in itself, but the question becomes: can these models answer not only *how*-, but also *why*-questions? Or, to put this differently, is it warranted to have a DST-C model provide a description of a system's dynamics, and suppose that that quantitative *description* constitutes a

qualitative *explanation*? I do believe that DST-C, and especially carefully-wrought models such as the one by Thelen et al. (2001), can help us travel at least part of the distance, but many questions remain. A more focused bit of criticism, ancillary to the main point, is one of the problems specific to Thelen et al.'s (2001) model of the A-not-B-error: what is the *h-parameter*? It is fine to consider it an abstract component of a model that provides a structural *description* of the system's behaviour, but, given the importance afforded to it by the authors, what does it *explain*?

My suggestion is that the holistic descriptions of behaviour of DST-C, and many forms of $E_{(i)}C$, leave a lot unsaid that needs to be investigated. Part of the reason for this book is to fill in some of those blanks.

3.5 - Previewing The Radicality Manifold: Interacting Domains

The most useful aspect of Thelen at al.'s brand of DST-C to my own project is the fact it offers a way to provide $E_{(i)}C$-compatible descriptions of behaviour, via their movement planning field. It will be my contention that we need to integrate additional kinds of processes and properties into the model in a more explicit sense if we are to *explain cognition* instead of merely describing behaviour.

Via the discussion of the ecology of colour perception (in section 4.5), I will investigate what the balance between the physical properties of the agent's body (e.g. the specifications of his retinal receptors) on the one hand, and the physical properties of his characteristic niche (the environment) on the other needs to be like to enable successful colour-based interaction (involving both perception and cognition) of the agent with that environment. My analysis of the linguistic categorization of perceptual colour space (in section 4.3) will yield a clearer idea of how the biomechanical properties of the agent interact with or are influenced by socioculturally shaped environmental structures (e.g. the way in which a specific language encodes colour categories). Then, it will prove to be a relatively intuitive step from linguistic categories to *conceptual* categories, and concepts proper.

Hence, following this trajectory along these theories as outlined will yield specifications of all the domains that are of interest in accounting for $E_{(i)}C$: DST-C gives us a handy description of behaviour, the ecology of colour teaches us about distal and proximal physical properties, the linguistics of colour adds the sociocultural domain, and opens up an avenue towards an analysis of concepts, the 'vehicles' of higher cognition.

I suggest that properties and processes that can be described in terms of these separate domains *collectively realise* the behaviour (action, cognition and locution) of an embodied and embedded agent. The 'Radicality Manifold'-model I will develop later in this book will conceptualise these domains as *spaces* as extrapolations of Thelen et al.'s dynamical movement planning *field*, each of the processes or properties in these

spaces offering *affordances* to other processes, i.e. one process or property constraining or enabling one or more other processes. I realise that this sounds awfully vague and abstract at this stage; but then again, so do some of the usual desciptions and explanations that are offered by other supporters of $E_{(i)}C$. I would like to ask the reader to bear with me for the duration of this journey; after a set-up via the philosophy of colour perception, I intend to cash out my promises by delivering my theories of embodied/embedded concepts and cognition in the second half of the book (chapter 6 and onwards).

= =

[SUMMARY of chapter 3 AND PREVIEW]

Using the tools of Dynamical Systems Theory, Thelen et al. (2001) have constructed a dynamical movement planning field, a model which expresses activity that is congruent with the basic behavioural choices made by young children when they fall prey to the A-not-B-error. This model is enactive in character, abolishing internal representation in its explanation of this minded behaviour. In chapter 7, I will take a closer look at an important related problem: the status of representation.

One of the problems of Thelen et al.'s model is that it pays little attention to more advanced cognition. Hence, this model offers a basic description of $E_{(i)}C$ behaviour in terms of a field-based model; an important task to be carried out in this book will be to expand and adapt that model for an explanation of concept-involving behaviour (see chapter 8 for the way in which this behaviour-based model is integrated into a broader $E_{(i)}C$ model). This adaptation will also help sidestep Newton's curse, which is the tendency to ignore contextual effects when extrapolating quantitative methodology into qualitative ontology. That is, the 'Radicality Manifold'-model to be developed throughout this book will provide a more elaborate account of the interaction of agent, physical world, social world, and concepts as a way to describe agentive behaviour. This expansion is multi-layered:
(1) expansion of the description of physical/ecological embeddedness;
(2) expansion into social embeddedness;
(3) a more dedicated description of the structure of this field (expanded into a higher-dimensional *space*), as pertaining to a description of concepts.

Hence, in the chapters to come, this expansion will occur in stages: the first clues about sensorimotor situatedness (agent-physical world interaction), and social situatedness will be provided in chapter 4, when these interaction modes as pertaining to colour perception are discussed. The idea here is that in theories of colour perception, there has been much attention for the ways in which basic sensorimotor contingencies help guide and inform more complex behaviour. Clues about the properties of this space extrapolated from this colour-focused approach will be given in chapters 5 and 6.

[4 - Agent-Environment Interaction: Ecology and Language of Colour]

4.1 - Two Sets of Theories

In the sections to follow, I will compare and contrast two sets of theories, to try and lay the foundation of an $E_{(i)}C$-appropriate account of concepts. The first set of theories (to be discussed in section 4.3) concerns the linguistic anthropology of colour, and my attempt to synthesize a workable account from the comparison of the universalistic and relativistic camps that emerged in the wake of the publication of Berlin and Kay (1969); this discussion will contribute a sense of *socio-cultural interaction* (i.e. enculturedness, social concepts) to my own endeavour. The second set of theories to be compared (in section 4.5), the ecological colour perception theories of Evan Thompson and Roger Shepard, mainly concerns the topic of *agent-environment interaction* (i.e. embodiment, embeddedness and enaction). I will modify and utilise many of the ideas present in this account of colour perception and phenomenology, because it represents the interlocking and interacting of many abilities, properties and processes that are also highly relevant to an account of concepts, and presents this interaction in a way that makes it an ideal, compact case study in preparation of a more comprehensive theory about $E_{(i)}C$: perceptually guided interaction (i.e. enaction) of an embodied agent with its physical environment (i.e. embeddedness), as well as conceptually guided interaction of an embodied and embedded agent with other agents (i.e. enculturedness).

There is one thing about the focus of the upcoming discussion that is absolutely essential to mention in advance. In the following sections, there will be clues about the content of the concept 'colour' - that is, I will develop certain ideas about what the *scientific* concept 'colour' is supposed to mean. While that is an interesting and worthwhile line of investigation on its own, the main purpose of this discussion is to derive from that discussion claims about the structural properties of concepts as such, i.e. what the properties of the concept 'concept' are. The phenomenon colour offers a good vehicle for such a line of investigation, because, firstly, both the embodied (an agent's sensorimotor contingencies) and embedded (the situatedness of an agent in his ecological *and* his social niche) aspects of said phenomenon have been well-investigated. Secondly, colour perception is almost never merely sensorimotor interaction: many colour-involving behaviour also involves more advanced judgments about the world. In other words: the way 'up' from basic perception towards concepts and cognition is comparatively easily made in examples involving colour. In that sense, the case study in this section and the next is as much about 'colour cognition' as it is about colour perception. Lastly, the rich discussion about the properties of phenomenal and/or linguistic colour space (see section 4.3 in particular) provides interesting clues about the structural properties of concepts which, as has been mentioned before, is intended to provide an addition to the dynamical field-based description of behaviour courtesy of

Thelen et al. (2001). The end result will be a new way of speaking about concepts and their properties, which is the goal of this whole excercise.

4.2 - Colour Phenomenology: The Received View

The received view of colour perception includes, as a significant component, the views of Ewald Hering. In his 'Grundzüge der Lehre vom Lichtsinn' (1920), Hering presented a phenomenological account of colour vision, claiming there are four primary chromatic hues: red, green, blue and yellow. Each of these is pure and unitary, whereas other chromatic colours can be introspectively understood to be an intermediary between two of these four, or some mixture. Orange, purple, chartreuse (yellow-green) and turquoise (blue-green) are secondary colours. Furthermore, he noted there appear to be certain constraints on the kinds of colour mixtures or combinations we can perceive: red and yellow or green and blue can peacefully coexist, but a combination or co-manifestation of red and green (as in a reddish green), or yellow and blue (as in a yellowish blue) appear impossible. Hering hypothesized (but was unable to prove at that time) that these opponency relations of the two pairs red vs. green and yellow vs. blue were somehow generated by some neural process.

At first glance, Hering's conclusions and knowledge glanced from investigations into the anatomy of the eye appeared contradictory: the human retina contains three types of cone receptors (i.e. humans are 'trichromats'), each most sensitive to a particular wavelength band. One cone type's pigment was shown to maximally absorb light with a wavelength of 445 nm, the second exhibited maximum absorption at 535 nm, the third at 570 nm. Intially dubbed 'blue', 'green' and 'red' cones, respectively, the failure to match the absorption maxima with the mean wavelengths of the actual colours mentioned prompted the adoption of a new convention designating the cone types 'S' (short wave), 'M' (middle wave) and 'L' (long wave).

In the nineteenth century, it was supposed colour sensations were generated by adding responses from the various cone types. However, if that were the case, the output at the retinal level would not differentiate between stimulus intensity and wavelength. Individual cones are not able to specify a particular colour sensation: a particular receptor will not be able to distinguish between a light at some specified intensity and wavelength and a light at double the intensity of a different wavelength for which this receptor is half as sensitive. Adding the response curves for more than one type of receptor will not help to alleviate this discrimination problem - one of the hindrances is that the cone types each respond to a fairly broad band of wavelengths, and these ranges show considerable overlap for the three cone types.

The step from three cone types to four chromatic colours in two opponent relations (and an additional achromatic opponency, black vs. white) appeared difficult to make. However, experiments conducted in the 1950s

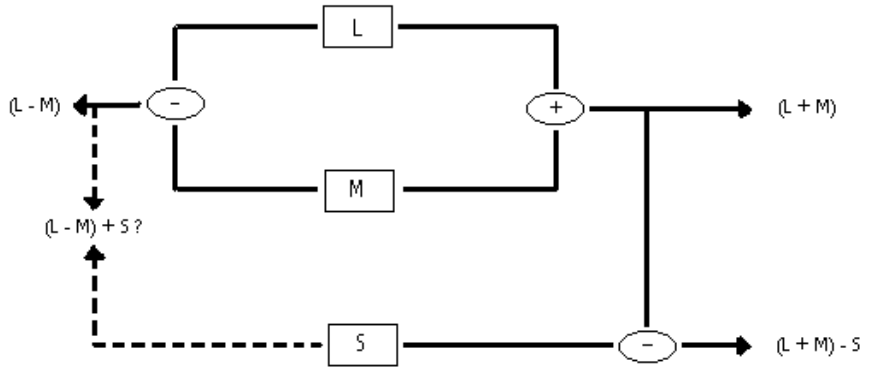
by Leo Hurvich and Dorothea Jameson appeared to quantitatively corroborate Hering's qualitative theory. Test subjects were given a light stimulus, and were asked to find the hue of a second light that would 'cancel' the hue of the first light. Phenomena corroborating Hering's opponency relations were indeed found[NOTE 16].

From this project a theory emerged that accomodated both the trichromacy of the retina and the four-hue opponency at the phenomenal level. Central to this theory, which in its general form (there is dissent at the micro-level) has become the received view, is the notion that retinal responses are not added, but compared. If the differences in receptor output are calculated, it turns out the discriminatory potential of the system is greatly enhanced - the response to a higher intensity at some wavelength no longer suffices to 'simulate' any response to a different wavelength.

These comparisons are initially established by a simple opponency-type mechanism in the neurons directly involved in the initial stages of retinal receptor response processing. These neurons enhance the differences between the signals by responding in some way (say, enhancing the signal) to stimulation of the centre of their receptive field (some region of the retina they receive and process signals from), but responding in a directly opposite fashion to the same kind of stimulation of the edges of their receptive field. Of these neurons, the 'on-cells' react excititory to stimuli reaching the centre of a receptive field, and inhibitory to the same stimuli at the edges of their receptive fields, while the 'off-cells' exhibit the opposite behaviour, responding positively to the removal of a particular stimulus from their receptive centre. As such, these neurons embody a spatially antagonistic 'push-pull'-system (Thompson, 1995) responding to fluctuations in the light stimuli, causing the visual system as a whole to exhibit sensitivity to contrasts rather than absolute intensities.

Visual neurons which integrate signals from different cone types generate a similar phenomenon, but now in the spectral domain, sensitising the visual system to differences between wavelengths. These single opponent neurons have been found in all organisms with colour vision. Double opponent neurons can compare the outputs of single opponent neurons by exhibiting opponent responses to different cone types (say, excitation by signals from L-cones but inhibition from M cones in the centre of its receptive field, and the inverse reaction at the edges of the receptive field), and as such will generate maximum output in areas exhibiting great colour contrast.

Hardin (1988/1993) provides one version of the abstract model coding the performance of functionally defined chromatic and achromatic channels, as a quantification of opponency theory. There are three (functionally defined) channels (depicted in figure 2): the chromatic red-green and yellow-blue channels, and the achromatic black-white channel, each defined by specific additions and/or subtractions of the signals received from each of the three retinal receptor types S, M and L.



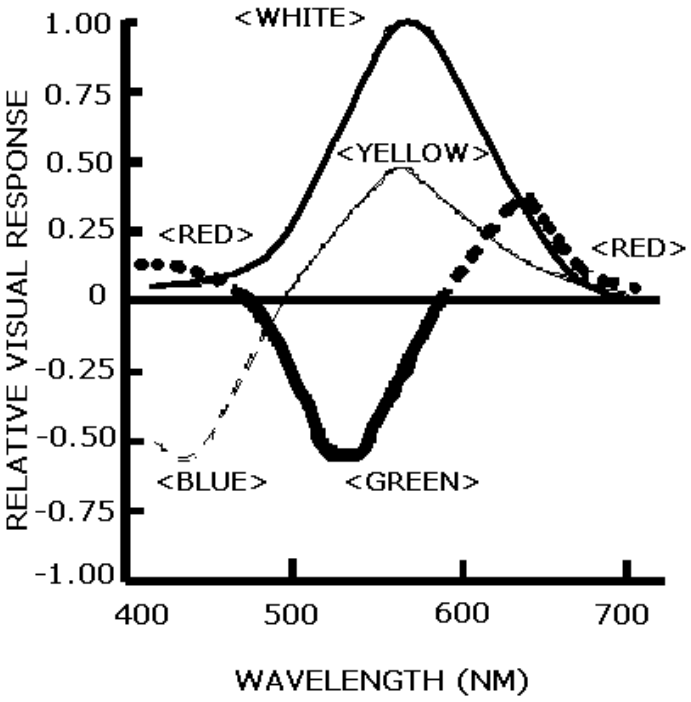
[Figure 2: A schematic depiction of quantitative opponent theory, adapted from Hardin (1988/1993, pg. 34)]

If '0' is understood to be the neural base rate (the activity at rest, resulting in the 'colour signal' dubbed 'brain gray'), Hardin specifies the output of the system might be catalogued as follows:

(L + M): achromatic channel
    (L + M) > 0 --> whiteness
    (L + M) < 0 --> blackness
(L - M): red-green channel
    (L - M) > 0 --> redness
    (L - M) < 0 --> greenness
(L + M) - S: yellow-blue channel
    (L + M) - S > 0 --> yellowness
    (L + M) - S < 0 --> blueness

The simple addition of signals from L and M receptors does not result in spectral opponency, and only codes changes in light intensity. It bears repeating these 'channels' are defined in *functional* terms - specifiying the exact realisation of these opponent processes is a matter of neurophysiological investigation, and most likely a very complex affair involving many different types of neurons and neural processes.

The phenomenal properties of opponent primary hues might be derived from this model (although in an experimental setting, the procedure is usually the other way around, with, for instance, Hurvich and Jameson's hue cancellation tasks providing the initial data). The performance of each channel (red-green or yellow-blue) might be expressed as a curve in a graph plotting wavelength in nanometres (*x*-axis) against the relative visual response, expressing the absorption spectra of the various receptor types relative to eachother (i.e. as processed by the neural channels), and recast against a logarithmic scale (*y*-axis). Figure 3 shows these curves.



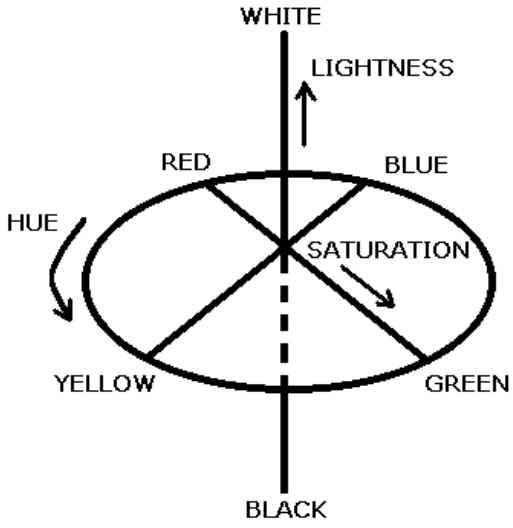
[Figure 3: Relative response functions for the opponent 'colour channels' and the achromatic channel]

Depicted this way it becomes easy to see why, for instance, red and green cannot co-exist at the same time at the same location (on some object's surface): if the red-green channel descends beneath the zero rate (the positive or negative character of some reponse merely signifying a particular attribution of coefficients, convenient within the model but ultimately arbitrary) for some part of the wavelength range, the resultant signal will specify a green-experience devoid of redness - other parts of the spectrum will result in a red-experience-specifying signal from that channel. Similarly, the yellow-blue channel will generate a yellow-signal if the curve rises above the zero rate, a blue signal if it dips below. Where the curve for one of the chromatic channels crosses the zero rate level (i.e. takes on the value 0), the only chromatic information provided comes from the other channel - either curve spans the entire visible spectrum. The achromatic channel is only specified by the whiteness response, since a blackness response can only be generated by contrast, for instance in perceiving a dimly lit spot surrounded by a region of high illumination. The saturation of a particular hue is specified as the ratio of the chromatic response to the sum of the chromatic and achromatic responses.

There are several different systems available to classify colours, but one of the most ubiquitous is the *Munsell system* (or systems based on the one developed by Munsell). The Munsell system organises basic characteristics

of the full range of colours of surfaces on three scales *hue*, *value* and *chroma* (the names Munsell himself specified), in a manner that lines up with the opponent structure as specified by the received view concerning colour phenomenology as described so far.

On this combined account, it is claimed that all perceived colours can be defined in terms of three parameters: *hue* (red, green, yellow, blue, purple or some intermediary), *lightness* (the measure of light reflection, ranging from ideal black - no reflection - to ideal white - maximum reflection) and *saturation* (the measure of distance from the gray of some lightness value, which embodies zero saturation). This results in a three-dimensional depiction of the way the various colour shades are related to eachother (expressing, for instance, the regularities that emerge in colour categorization tasks), called *perceptual colour space* (see figure 4).



[Figure 4: perceptual colour space]

Representing the three determinants of colour in three-dimensional space, lightness defines the vertical axis with black (0) at the nadir and white (10) at the zenith; hue is defined orthogonal to lightness along the perimeter of a colour circle divided into ten regions (red (R), red/yellow (RY), yellow (Y), yellow/green (YG), green (G), green/blue (GB), blue (B), blue/purple (BP), purple (P), purple/red (PR)) with each region divided into ten steps; saturation is a measure of distance from the origin (the achromatic centre, halfway between black and white and of neutral hue) divided into twenty steps. On each dimension, the distance between steps is intended to represent equal perceptual difference in normal daylight, in a neutral (i.e. gray to white) environment. Since not all hues are represented at all levels of saturation or lightness (most typical yellows are of relatively high lightness but low saturation, whereas most typical reds are of relatively low lightness but high saturation), the resultant three-dimensional structure is not a perfect sphere. The Munsell notation specifies a colour by giving the

appropriate values for each of the three scales thus: [hue] [lightness]/[saturation]. 7GB 4/6, for example, identifies a shade of turquoise.

4.3 - Sociocultural Situatedness: The Linguistical Anthropology of Colour

In linguistics and anthropology, there have been discussions about the way perceptual colour space has been used to characterise human colour perception, and in particular, how people from different cultures carve up this space with colour terms. The term 'red' would indicate a particular region in perceptual colour space, with the best example of red forming the focal point of that broader region of red and reddish shades; the same for 'yellow', 'green', 'blue' and so on.

Berlin and Kay (1969) stated that languages can be ordered in accordance with the number of basic colour terms they contain, in the following evolutionary sequence:

[white / black] < [red] < [green / yellow] < [blue] < [brown] < [purple / pink / orange / gray].

So, this sequence depicts the evolutionary order of the lexical segmentation of perceptual colour space. This means that the most primitive languages, in terms of colour lexicon, will only have two colour words, with English glosses 'white' and 'black'. If a language has a somewhat more complex lexicon with three basic colour terms, the third word will always be a term denoting red, for an even more complex language the fourth term will deote either green or yellow, and so on.

For their experiments, Berlin and Kay used a two-dimensional array of colour chips of varying hue and lightness, all at the maximum available saturation - this was basically (some subsection of) the peel of the complete three-dimensional structure depicted above. The sometimes-utilised Farnsworth-Munsell array also portrays colours of varying hues and lightness, but at a much lower level of saturation.

Kay (1975) provided the following refinement of the evolutionary sequence:

Stage I:[WHITE *and* BLACK] --> **Stage II**:[RED] --> **Stage IIIa/IV**:[GRUE > yellow] *or* **Stage IIIb/IV**:[yellow > GRUE] --> **Stage V**:[green *and* blue] --> **Stage VI**:[brown] --> **Stage VII**:[purple, pink, gray *and/or* orange].

'Grue' is a combined green/blue category. Berlin and Kay state that the colour perception of humans from all cultures is structured in the same way, i.e. that the respective colour regions' foci lie at the same general locations in perceptual colour space, and that the differences in colour language across cultures have little to no influence on actual colour *perception*. This claim amounts to a rejection of the so-called *Sapir-Whorf thesis* regarding the categorization of perceptual colour space. The Sapir-Whorf thesis is a

notorious theoretical claim, expressing the idea that there is a co-dependency of language, thought and culture. In other words, the thesis says an agent's socio-cultural environment is of (at least non-trivial) influence on his cognitive (and also perceptual) processes. Berlin and Kay defend a universalism regarding the biological structure involved in generating colour experience: they say that despite differences in colour language, colour experience (including the structure of perceptual colour space depicted above) is the same for all humans, because they share the same sensory processing mechanisms.

Apart from a great deal of critical acclaim, rising to a point of revered prominence in that the theories professed by and built upon the work of Berlin and Kay have grown to be the received view on this crossroads of linguistics, neurophysiology, philosophy and anthropology, there is a reasonably strong undertow evoked by critics dismissing the universalist tradition's central tenets. John Lucy (1992) develops a sustained critique of the research of Berlin and Kay, and one of his most important ideas concerns the *artificial nature* of said research. That is, he charges Berlin and Kay with the accusation that their approach to colour language is decontextualised: any references to the socio-cultural context or the natural way of functioning of colour language are avoided.

This decontextualisation occurs in three ways:
-(1) due to the methodological secession employed towards studying colour terms in abstracta, a failure to properly assess the grammatical properties of colour words disables a clear view on the structural divergences of the various languages studied.
-(2) a similar abstraction is achieved in the perceptual domain: colour is studied as a perceptual primitive, its connections to any habitually co-occurring experiential or environmental features severed. Any natural synchrony between a colour and some property of whatever it is colours are thought to be a significant feature of in a particular language community, even cross-cultural regularities of such a kind, will not be found.
-(3) any semantical features of colour words are reduced to mere denotation by asking test subjects to do no more than link a decontextualised hue stimulus to a basic colour term.

Summarising: the very idea Berlin and Kay attempted to investigate, the Sapir-Whorf thesis, predicts an influence of the language one uses on the way one interprets reality, how one constructs a conceptual world-view. However, in much colour term research, this broad thesis was constrained to one merely involving the processing of decontextualised perceptual information. Lucy's suggestion is, then, that Berlin and Kay's research did not discount the Sapir-Whorf-thesis because their experimental setup was such that any supporting evidence could never be found: the cultural-linguistic context that forms the very core of the thesis had been abstracted away. Lucy supports the opposite position: he is a relativist concerning the influence of language on colour experience. That is, he states that when your colour language is different, your colour experience is also different.

One clue that this can be so is that in many cultures, terms that can be (and often are, for instance by Berlin and Kay, 1969) given English colour words as glosses, are actually words that refer to much more than merely colour shades. A telling example, derived from Lucy (1997), states that Hanunóo (Philippines) colour words exhibit a wide referential range not restricted to mere hue-name correlations, or even to typical colour/utility connotations (as have been found in, for instance, certain African tribal languages, where some colour terms refer explicitly to cattle colours). In their colour language, Hanunóo speakers use four basic colour terms: lagti' (light), bi:ru (dark), rara' (dry) and latuy (wet), with English glosses - attributed by Berlin and Kay - white, black, red and green, respectively. However, the Hanunóo not only encode colorimetric information such as oppositions between light and dark, but also between dry and wet, or deep/unfading/desirable and pale/colourless/weak. All this additional referential information is not captured by the English glosses. This means that for the Hanunóo, and many other linguistic communities, words denoting colour-relevant information are necessarily linked to other properties.

Criticism like this has resulted in a rejection, by some, of the strong claim that the evolutionary order of colour term acquisition presented by Berlin and Kay is a universally occurring structure: the multi-referentiality of colour words between different languages and cultures means that the order in which colour words are learned depends on many contextual (i.e. environmental, socio-linguistic) factors apart from neuro-physiological structure. As a very general indication of evolution, the Berlin and Kay sequence might offer some useful suggestions, but many aberrant cases (such as the Hanunóo example described above) and the inordinate focus on abstract colour terms render this part of Berlin and Kay's thesis less universally applicable.

However, the psychophysical model described above (in section 4.2) is a fairly straightforward case of functionalistic reduction: the apparent conflict of a three-receptor retina with the phenomenal four-hue opponency is recast in functionalistic terms, after which an attempt at providing a neurophysiological description of the structure specified is carried out, yielding the desired physicalistic/universalistic explanation of the categorization of perceptual colour space. There have been some mildly successful steps towards completing this reduction, but not enough headway has been made to warrant proclaiming the operation an unqualified success. In brief: the three-way explanatory gap between the neurophysiological, phenomenal and linguistic levels remains. However, it does appear possible to isolate regularities on each of those levels.

Some of those regularities are uncovered by Kimberly Jameson and Roy G. D'Andrade[NOTE 17]. They defend an interesting hypotheses in their 'It's not really red, green, yellow, blue: an inquiry into perceptual color space' (1997). They claim irregularities in perceptual colour space facilitate its progressive compartmentalisation that turns out to line up fairly neatly with

Berlin and Kay's evolutionary sequence. And this can explain why red, green, yellow and blue appear to have obtained a special status, despite the fact there are difficulties in actually lining up the Hering-style opponency intuition with the available physiological data.

The Munsell-type three-dimensional colour solid (see section 4.2) is a structure quite unlike a perfect sphere, with protrusions at places where the saturation for particular hues can be specified at much higher levels (in the case of red, for instance), and indentations at places where saturation levels are unavailable beyond relatively low values (in the case of yellow). Because of these irregularities, distances between foci are not uniform. If the most primitive lexical segmentation of colour space has been made by separating colours into dark/cool and light/warm categories, the most informative additional term that might be acquired specifies RED, which has a focus farthest away from the regions specified by the intial two categories. The fourth most informative colour word to be acquired would be either yellow or blue, followed by green, purple, pink, orange, brown and gray (as determined by distance computations carried out by Boynton and Olsen (1987)). The evolutionary order thus generated does not diverge widely from the findings of the World Color Survey[NOTE 18], claim Jameson and D'Andrade.

If the above is coherent, the classical four-hue opponency account nor the three dimensional phenomenological model are appropriate in the form they are usually presented in, even on phenomenological grounds. Despite this, Jameson and D'Andrade do wish to accept the theoretical outlines of the three-dimensional phenomenal colour space and opponency, and merely suggest a different (and somewhat more complex) account of basic hues and their phenomenal properties. Specifically, they wish to deny the four-hue model and propose a five-hue segmentation. Still, jugments of similarity between Munsell colours will yield a particular structure in three-dimensional color manifold (of hue, brightness and saturation). The perceived spacing of hues, if one attempts to account for a wide variety of deviant experimental observational results, will indeed yield a non-Euclidian structure, but this will at least locally exhibit normal Euclidian metric properties - there is a substantial center region that exhibits high regularity.

The account provided by Jameson and D'Andrade, stating three-dimensional phenomenal colour space should incorporate a fifth hue category, compares reasonably favourably to Kay's (1975) modifications to Berlin and Kay (1969), in which he attempted to account for some discrepancies of the original theory with data found by Rosch (1972). As explained above, Jameson and D'Andrade suggested a process of colour lexicon expansion in which the most informative colour term (in terms of distance from the colour words already in use) would be the most logical next choice. In his (1975), Kay embraced the relevance of socio-cultural influences on the acquisition - by individual speakers or social subgroups - of colour terms beyond those utilised in a language residing at a particular evolutionary stage.

However, the causes for the universality of colour terms as such appear to be somewhat different from those present in the (opponency/physiology-based) model proposed by Kay and McDaniel (1978), in the sense that the modification of perceptual colour space (by including purple as a fifth axis-defining hue) necessitates an additional step in the explanatory chain from neuronal response categories to colour space lexicalisation. It remains to be seen to what extent socio-cultural factors might influence this additional step, involving the distance-based compartmentalisation of colour space as described above - it does not appear far-fetched to suppose a dominant colour in some ecological niche might be named earlier by this environment's inhabitants than some evolutionarily prior colour, despite a greater perceptual distance, merely because the informational potential of lexicalising that environmentally significant colour overrides strictly phenomenological considerations.

Despite the convergence with the claim by noted critics of Berlin and Kay, Saunders and Van Brakel (1997) that the categorization of colour space is not determined solely by phenomenal and neurophysiological factors, the above does yield the view that there are some non-trivial constraints on colour categorization - there appears to be some logic to the way the colour space can be incrementally lexicalised. Hubey (1997), examines various algebraic and geometric analyses of colour space and language, and reaches a similar conclusion.

In an idealised three-dimensional vector-space depiction of the interrelatedness of colours according to the *Commission Internationale de l'Eclairage* (CIE) diagram, the tristimulus colours red, green and blue form the nodes in a triangular plane orthogonal to the black-white axis. Black has coordinates '000' (i.e. resides at the origin), white '111', red '001', green '010' and blue '100'. Red, green and blue are the additive primaries (each reside on an axis of the vector space), the multiplicative (or subtractive) primaries are yellow (011), cyan (110) and magenta (101).

Representing the foci of Berlin and Kay's basic colour terms unveils an interesting phenomenon: the distribution is asymmetric in the sense that the majority of them appear to reside near the red and green nodes. Adapting for the heightened sensitivity of the human visual system for these high-saturation hues would entail extending the red and green axes accordingly. With the vector space thus modified the scenario suggested by Jameson and D'Andrade is corroborated: after the division of colour space via the lexicalisation of black and white, red (the longest wavelength, to which the visual system is more sensitive) is farthest away from either, thus is the prime candidate for intial hue lexicalisation. After this, there is a choice between either green or blue - green wins out, once more because the human visual system favours the longer wavelengths. The prominence of yellow at this stage might be due to its abundance in nature (the fall leaves, or dry grass); after yellow, blue is lexicalised. The largest 'open' region still remaining resides in the red section of colour space, which is probably why

most of the additional colour terms pick out hues near the red node - gray (along the black-white axis, at the center) and purple (nearer blue) are the exceptions.

It deserves noting that in the model above, there is a fairly elegant account of opponency, as well as a satisfying explanation for the status of yellow as a unique hue, despite it not being an additive primary. These two explanations are related, and have to do with averaging effects. Just like the addition of all three primaries will yield 111, i.e. white, averaging red and green will yield not reddish green, but yellow. Yellow is the only multiplicative primary that stands out as a unique hue, whereas magenta and cyan do not - this is due to the relative wavelength-wise proximity of red and green, addition resulting in a much higher, amplified peak at the yellow wavelength than when, red and blue are added. The red and blue wavelength ranges are further apart and as such show much less overlap, resulting in a saddle node at magenta perceived as a mixture of red and blue.

Whether the hue foci themselves buttress the above-mentioned process of progressive segmentation of the colour space, or a sequence of increasingly fine-tuned delineations of broad categories as suggested by Roberson, Davidoff and Davies (2000) cannot be stated with any semblance of certainty at this point. A claim that does appear at least tenable, and which I would indeed wish to defend, is that socio-cultural and linguistic factors might have a non-trivial influence on the way this process bears out.

So, socio-cultural influences (linguistic ones in particular) are not to be ignored in explaining the acquisition of an appropriate colour sense - but neither are biogenetic factors. The intuition is that both the human neurophysiological make-up, and their socio-cultural embeddedness (chiefly via linguistic interaction) exert non-trivial influence on the way one's behaviour regarding colour is given shape.

There are certain constraints on the range of options open to human beings in constructing methods, customs and intra-culturally dispersed habits that help achieve some goal (such as establishing a shared colour lexicon). Sociocultural peculiarities, assumed to be of at least some influence on colour categorization, may explain moderate differences between colour language between cultures, but this does not automatically negate any and all laws describing some universal order - there is little dispute that the biological composition of the colour vision system and the performance characteristics are generally the same (discounting deficiencies) in all humans - but said order might find different expressions in various cultural contexts.

So it seems that the accounts involving opponent hues and the three-dimensional characterisation of colour space (defined by hue, saturation and brightness) do not tell the complete story - the contextual information

encapsulated in the broad referential range of many language's colour terms cannot be cast aside as callously as Berlin and Kay's basicness criteria would have us do - but it does tell at least part of it. The success of the Hurvich and D. Jameson cancellation experiments (see section 4.2) and the kinds of opponency found in various types of neural cells suggest that at certain physiological as well as phenomenal levels, some kind of opponency is indeed occurent. However, an accurate account will in all likelihood be rather more complex than the model currently in wide use.

Despite the fact that the psychophysical model of colour perception is incomplete (the $E_{(i)}C$-appropriate *context* is mostly missing), I suggest that the basic idea behind the model *as a way of structuring a subclass of behavioural responses* as a result of the embodied processing of stimuli is sound. For instance, it is possible to develop a similar account of perceptions in other modalities. The possible spectrum of perceived sounds, for instance, can be defined in terms of the parameters *amplitude*, *pitch* and *compactness* (Jakobson and Halle, 1956), like so:

((Dimension))		(sound)	:	(colour)
total energy	-	amplitude	:	brightness
frequency	-	pitch	:	hue
purity	-	compactness	:	saturation

The developmental sequence Jakobson and Halle uncovered for sound is ontogenetic, referring to the increasing sophistication of phonological distinctions as infants develop, whereas the colour sequence Berlin and Kay's theory yields is phylogenetic: an evolutionary scale (which, it needs to be noted, is of *cultural* rather than biological inclination).

In both the case of phonological development and the evolution of colour space categorization, the initial stage contains two categories: the first uttering of an infant is /pa/, uniting the contrasting diffuse stop /p/ (closest to silence; minimal energy) and open vowel /a/ (loudest; highest energy); in the case of colour, black (minimal energy/brightness) and white (highest energy/brightness) are the two initial categories.

The first colour term to be acquired after this in both sound and colour consists of an exploration of the frequency dimension, and in both cases in the low end of the energy spectrum: whereas /p/ exhibits low tonality at low loudness, the acquisition of /t/ exhibits high tonality, but still at low loudness. In the case of colour, the low-brightness hue RED joins with BLACK to contrast with WHITE on the brightness scale, but joins with WHITE to contrast with BLACK on the hue scale.

Hence, both sound and colour exhibit a increasing sophistication in categorising phenomenal space as the child (for sound) or the culture (for colour) 'matures'. Berlin and Kay briefly hypothesize that colour language might have foundations similar to syntax and phonology, namely that some 'species-specific bio-morphological structure' (Berlin and Kay 1969, pg. 109)

determines the particular evolutionary order that was found. These parametrizations - of colour in terms of hue, brightness and saturation, and sound in terms of amplitude, pitch and compactness - yield a series of *perceptual spaces*.

4.4 - Towards Contextualised Concepts

Now I can offer a somewhat more concrete answer to the question why I discuss colour, when I should be talking about concepts. This idea of perceptual spaces is why: I will use them to develop the idea of *conceptual* spaces, to put a bit more meat on the bones of a Thelen et al.-based, $E_{(i)}C$-appropriate theory of concepts. In order to prepare for a more detailed specification of the properties of such spaces, chapter 5 explores the properties of the aforementioned perceptual spaces, focusing on the case of colour space. After that, a more general account of conceptual spaces will be built from that colour-related prototype.

But first, as stated above, there is one major issue that needs to be stressed: Lucy's worry that an account such as the one above is highly decontextualised. In other words: the universalism of the received view needs a little relativist tweaking. A prudent move to address this issue would be to state that *both* universalists and relativists have some suggestions of value to make: the properties of the the human sensory organs and influences from an agent's specific socio-cultural and linguistic environment *collectively* specify said agent's colour-related behaviour.

However, one aspect is still missing from this description: the influence of the *physical* environment. Enactivism ($E_{(A)}C$) offers the tools to specify the role of that particular aspect: as was said in section 1.2, enactivism attempts to define cognition and perception in terms of the ways in which an agent interacts with his environment. Evan Thompson, one of the main defenders of enactivism, provides such a story for colour perception. I will show that this interaction involves a co-attunement of agent and environment in a way that can, in part, be described in terms of the 'conceptual spaces'-account hinted at above. I will use clues about ecological theories of colour perception, to be described in the next sections, to get a clearer picture of the character of that co-attunement. Furthermore, my own project of constructing an $E_{(i)}C$-appropriate account of concepts will contain descriptions of a contextualised, rather than a decontextualised, apprehension of concepts - the basic mistake made by Berlin and Kay as described by Lucy is the one that I will have the explicit goal of avoiding.

4.5 - Physical Situatedness: The Ecology of Colour

4.5.1 - Colour Enactivism

Evan Thompson (in Thompson, Palacios, Varela 1992, as well as Thompson 1995a, 1995b, 2000) argues for a *relational* account of colour

perception, with significant influences from ecology. The biological function of colour vision, he claims, is not to retrieve information about any single type of physical structure (for instance the microphysical surface structure of an object, with specific colour reflectance properties), but rather to help guide an organism's actions in his specific ecological niche. This implies that colour is not a property to be found either in or on the perceived object, nor in the perceiver's mind, but a fundamentally context-dependent action-guiding principle, i.e. something that emerges in the *relation* between organism and environment.

Thompson explicitly constructs his argumentation in support of his own ecological approach to colour vision as a response to the constraints of what he feels is the cul-de-sac in which the regular colour theories now find themselves. Explanations of the processes yielding colour vision in the received view, which Thompson says trace back to Locke, customarily involve the following components:

(1) the physical structure of the object, anchouring the disposition of said object to reflect light of a particular kind;
(2) the composition and properties of the light, and its disposition to affect the perceiver's senses in a specific way;
(3) the perceiver's colour sensations, which are explained by processes involving the two components above.

Such a basic structure, Thompson claims, underlies both the modern subjectivist and objectivist theories of colour - the subjectivist will claim colour is to be identified with some aspect of the perceiver's sensation, whereas the objectivist will wish to define colour in terms of properties of the distal object. Whatever component is accentuated, Thompson says the overarching framework of the received view is one expressing a representationalist inclination.

Thompson isolates the tension between objectivism and subjectivism about colour to be (one of) the main philosophical problem(s) pertaining to colour vision. Within the basic model described above (object, perceived to be coloured + light + subject, doing the perceiving), there is a lack of consensus about which component deserves ontological primacy. The two extremes of what is a continuum containing many different intermediate positions, are objectivism and subjectivism.

-*objectivism*: colour is a property of objects, for instance surface spectral reflectance[NOTE 19]. This position is defended by Alex Byrne and David Hilbert: 'physicalism - reflectance physicalism, in particular - has the resources to deal with common objections, and can be smoothly integrated with much empirical work' (Byrne and Hilbert 2003)

-*subjectivism*: colour is an aspect of or generated in a perceiver's experience. A rather specific version of this position (eliminativist reductionism) is held by C.L. Hardin: 'We are to be eliminativists with

respect to color as a property of objects, but reductivists with respect to color experiences' (Hardin 1988)

The problem is that neither objectivist theories, such as physicalism or dispositionalism, nor subjectivist theories are capable of telling the whole story about what colour is[NOTE 20]. Rather, each of the various positions appears to possess *part* of the puzzle: they each have their own subsection of what colour appears to be that they can explain best, but each position also has major shortcomings outside of their privileged explanatory region.

Evan Thompson's alternate suggestion is to construct an enactivist (i.e. $E_{(A)}C$), relational, *ecological* theory about colour. His theory is ecological because it attempts to explain colour as an aspect of the way a person stands in specific relations to his environment, and he intends this to be an attempt to progress beyond the objectivism-versus-subjectivism-debate. Thompson's very general (and hardly controversial) initial hypothesis about colour is, that:

'color vision generates a relatively stable set of perceptual categories that can facilitate object identification and guide behaviour accordingly' (Thompson 2000).

At the core of his way of finding a theory that can both align with this general guideline *and* transcend the standard objectivism vs. subjectivism dichotomy, Thompson's approach contains the thesis that colour is a *relational* property:

'The basic thesis of the ecological view is that colours are properties that depend on both colour perceivers and their environments. Colours are not intrinsic to objects in the physical world (computational objectivism) or to neural processes in the visual system (neurophysiological subjectivism); rather, they are properties of the world taken in relation to the perceiver. Thus on the question of whether colours are intrinsic properties or relational properties, I side with the received view that they are relational. But unlike the received view, the relational position that I shall defend is distinctly *ecological* in the sense pioneered by the psychologist J. J. Gibson.' (Thompson 1995a, p. 177)

The relational approach that Thompson supports is not new, and in fact exhibits close resemblance to standard dispositionalism (the object has a disposition to cause a colour experience of a particular kind in a suitable perceiver). As the quote above shows, Thompson adds the clause that an ontological account of colour should also be *ecological*.

The term 'ecological' the way Thompson uses it has been imbued with meaning comprised of three elements: (1) Gibson's ecological vision, (2) comparative biology and (3) naturalism.

(1) *Gibson's ecological vision*: Thompson's moderate dismissal of the theories of colour of the received view - which in some way, shape or form involve the internal representation of some external state of affairs - is grounded in his enactive approach to cognition, a relational and action-oriented conceptual inclination with firm roots in J.J. Gibson's 'Ecological Approach To Visual Perception' (1979).

The basic argument Gibson develops contains the notion that a representationalist view of human cognition is inadequate - that is, a static abstraction in which a subject perceives (some aspect of) an object and forms an internal representation of said object after which a strategy might be devised to utilise or avoid this object, fundamentally misconceives the dynamic character of the agent/world-interaction. The perception of something in the environment is direct and unmediated, and in its very essence geared towards existing in the coherence of possibilities and constraints instantiated in the ecological dynamic. Perception, on this account, is not in its basic form a stationary affair, but occurs in ambient or even ambulatory situations in which the way the environment appears changes constantly, and the knowledge distilled from this dynamic converges on invariant-extraction, the awareness of constants and their interrelatedness from the ambient optic array.

A central concept coined by Gibson is 'affordance'[NOTE 21], that constellation of possibilities for action evoked by an object or organism in its coherence with the perceiver. A tree, to a bird, would offer the affordances of a source of fruit to eat, a place to build a nest, a place to hide from larger birds that might prey on it, and so on, while the same tree would posit to a human the affordance of shade from the sun, or fuel for his fires, and perhaps also fruit to assuage his hunger. Perception, then, is not simply the passive reception of sensory stimuli, but the active appraisal of possibilities for action - the world is not a Cartesian res extensa filled with extended things one might decide to use or not, but the layout of the environment and everything in it are already significant to the perceiver. This means some type of inference towards an apprension of usefulness from the impressions the perceiver collects is not needed - this is exactly the kind of internal representation Gibson argues against.

So, representationalism, according to Gibson's approach, fails to provide an adequate account of perceptual processes due to the fact it requires information coming from the objects to be modified, for instance by 'filling-in' mechanisms contained within the sensory system. Here, there is an explanatory gap between the physics of the object and the physiology of the perceiver.

Direct perception, in contrast, merely posits a resonance of the sensory apparatus with information provided by the distal object, thereby avoiding addition or subtraction of information by the perceiver's senses. In Gibsonian theories, speaking of an explanatory gap between physiology and psychology is rendered erroneous akin to commiting a category

mistake. For his ecological account, Thompson tones down this result somewhat, but it does provide at least some of the thrust of his argument (which will be summarised below).

However, this latter gap does again pose a problem for the representationalist, and this schism is possibly bigger than the physics – physiology divide: at some point, there is supposed to be a transformation from a physical and objective stimulus into a mental and subjective percept, and this change involves the attribution of meaning to intrinsically meaningless signals. In the theory of direct perception, an affordance (as the ecological object of an organism's perception) simply is defined as intrinsically expressing meaning and utility for that organism.

Summarising, the Gibsonian component of Thompson's theory of colour states there is a co-attunement of animal and environment. The animal is understood to be an *active* explorer rather than a decontextualised stimulus-receptor, and the world is not merely a space filled with objects, but a lived-in environment. This means that Thompson's theory of colour perception aligns most explicitly with theories involving situatedness (colour emerges due to the interaction of animal and world, and their respective properties) and enaction (colour perception is an active process of exploration).

Fodor and Pylyshyn (1981), interpret the ecological theory of perception as wishing to offer a theory not just of perception as such, but of cognitive processes in general (which would be a factor of importance in the interaction with affordances). They claim that, by its inordinate focus on perception, Gibson's project woefully underestimates the importance of constructing a representation-free theory of intentionality (the context in which the affordances would be illuminated, at the very least in the case of human perception) if one wishes to do away with representation as a factor in cognition. After all, they say, perceiving simpliciter is an extensional relation, whereas cognitive relations - seeing *as*, for instance - is intentional, and, on the standard account, as such requires some measure of mental representation to perform the interpretation-task. One can see an object, but this is a rather different state to be in than seeing said object *as signifying this or that*, or this object being in possession of some property or other, to an important extent because such a relation is dependent on one's idiosyncratic array of background knowledge, concerns, needs and wishes at a particular moment. Fodor and Pylyshyn wish to deny that the directness of perception extends to an apprehension of affordances, and stress the need for active (cognitive, representational) analysis involving intentionality in such cases. They do not claim it would be impossible to develop a theory of intentionality in which representation would be absent, they merely note Gibson does not offer a solution to the problem, which renders his theory incomplete, at the very least, and possibly fundamentally flawed in its overwhelming allocation of importance to the process of perception in itself.

Gibson would be able to object that on his account, perception cannot be severed from the action in which it is embedded - *on the contrary*, perception is defined as perceptually guided activity. Thompson explicitly endorses this view, on which it is claimed an organism does not merely perceive, after which this input is transferred to some internal processing unit, but acts and reacts (displays perceptuomotor adjustments) in a dynamic dance with his environment. This entails perception is something that does not take place in the brain, but a process that is instantiated in the organism as a whole, in the interplay of physical processes of perception and correlating movement, yielding a shift in vantage point from which a different array of possible perceptions emerges, necessitating some muscular response, and so on.

(2) *comparative biology*: Thompson's theory of colour is intended to apply not merely to humans, but to all livings being who can guide their behaviour based on chromatic perception. Because members of other species might have different types of eyes which are sensitive to other wavelength ranges than the human eye, or they might use their kind of colour vision to detect different properties of the distal scene than humans, this has consequences for the kind of property 'colour' can be claimed to be.

(3) *naturalism*: with his ecological account of colour, Thompson strives to capitalise upon the advantages afforded by the Gibsonian viewpoint (a bridge across the subject-object gap), but resting on a decidedly naturalistic foundation – that is, incorporating the latest knowledge gained in the relevant scientific fields. Therefore, he explicitly enlists the aid of scientific data from fields as diverse as neurophysiology, ethology, phenomenology and computational vision. With these tools, he hopes to transcend the subjective / objective-tension of the received view.

4.5.2 - The Evolutionary Adaptation to Illuminant Invariants

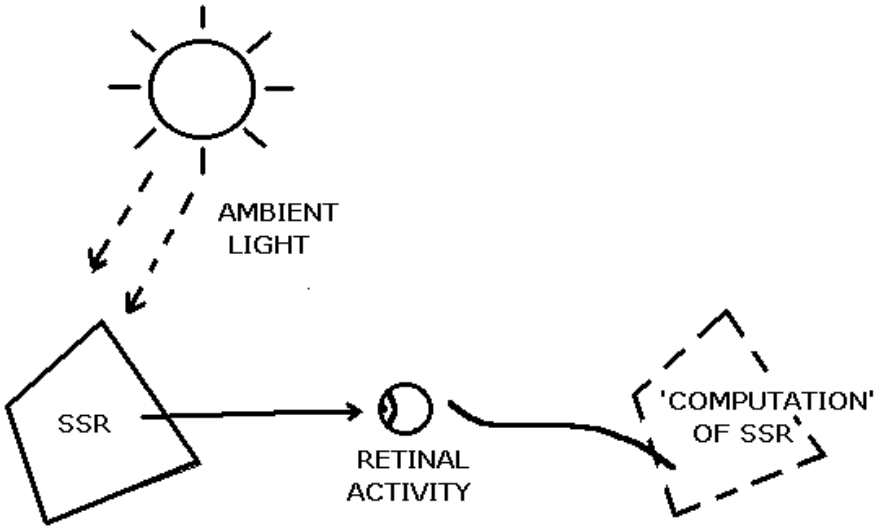
One of Evan Thompson's targets (in particular in Thompson 1995a and 1995b) is Roger Shepard, and his ecologically inspired theory pertaining to colour vision. This is interesting: Shepard, calls his theory about colour 'ecological', just like Thompson, but still they appear to suggest two diametrically opposed accounts of what colour is supposed to be. That is, Shepard suggests a computational approach, which is exactly the kind of view Thompson argues against. One objective of this section is to compare these two views.

But before that, I will offer a brief description of Shepard's ideas. Over the course of decades, Shepard has built a case for the co-evolution of animal and certain structural aspects of the environment, to yield what he calls 'perceptual-cognitive universals' (Shepard 1987, 2001). In the case of colour, his main claim is that, over the course of evolution, the chromatic vision system has become attuned to large-scale invariants in the optic array. Using the Linear Models Framework[NOTE 22] for support, he says

(human) trichromacy is a reflection of the three degrees of freedom of terrestrial illumination.

In broad terms, the account of colour vision which can be said to underlie Shepard's theory about the perceptual organisation of colours is the following: colour vision serves to gather stable and accurate information about the environment based on the chromatic composition of the light despite changes in luminance, and this information may be used to guide behaviour. If you deconstruct this description, you see all the familiar components of the received view as Thompson described it: there's a coloured object and a perceiving subject, the colour of the object provides information to the organism, and that information makes a behavioural difference (a food-coloured patch is something we might want to approach, a predator-coloured patch is not). An essential additional thesis is that the re-identification of some object based on colour requires an animal to exhibit some kind of *colour constancy*.

Computationalist approaches to explaining colour vision understand colour constancy to be a central feature of colour vision: it is the ability to identify some object as having a particular colour under a wide variety of illumination conditions (despite the light reflected by, say, a white object under a red lamp being essentially the same as a red[-dish] object under a white lamp). In functional terms, the purpose of colour vision would be to extract some invariant from the perceived scene - the physical structure identified to be that invariant is the object's surface spectral reflectance (see figure 5).



[Figure 5: Computational model of Colour Constancy (elimination of the effects of ambient light from the retinal input yields an estimate of the object's Surface Spectral Reflectance)]

The problem of explaining how the visual system would accomplish such a feat is the 'inverse optics'-problem. Thompson et al. (1992) describe the problem in its general form as: ‘the recovery of what are taken to be objective attributes of three-dimensional scenes from ambiguous two-dimensional projections‘.

In the case of colour vision, the problem takes the following form: "Because the retinal activity from a given point hopelessly confounds the illumination with the reflectance properties of surfaces, the core problem is to disentangle these variables and assign colours that correlate with surface properties. (...) In the case of color vision, the problem is to discard the source illuminant E and retain the invariant spectral reflectances ρ of object surfaces given only the retinal activity S" (Thompson et al. 1992).

Historically, an important strategy utilised in computational approaches to this underconstrained problem is to *introduce* constraints - in the case of colour vision, the main strategy is to use low-dimensional models of light and reflectances. An application of this idea is the aforementioned Linear Models Framework (LMF) - the basics of this model are described in Wandell (1989). His central thesis is that both ambient light and an object’s SSR can be described by linear models containing the weighted sums of a very limited number of basis functions (as explained in note 22).

With the spectral sensitivities of the chromatic receptors (for humans, the three types of cones), it is possible to formulate a reasonably simple equation expressing photoreceptor response as related to the distal surface properties and the illumination. A weighted sum of a limited number of basis functions suffices to describe all relevant variations of naturally occuring illuminants and SSR’s - perfect colour constancy can be achieved if the number of degrees of freedom of the reflectance equals the number of sensor types minus 1.

The basic setup for which this model attempts to specify the appropriate relations in terms of equations, consists of the components ambient lighting, an object with a specific SSR and the perceiver with specific chromatic receptor sensitivities. What the model needs to find is an estimate of the object’s reflectance properties.

Maloney (1992) claims that if certain assumptions are satisfied, LMF's algorithms will result in predictions of the object's surface properties, i.e. exhibit (perfect) colour constancy: any variations in ambient lighting conditions are 'filtered out' to yield information about the object's chromatic surface properties. One of the most important of these assumptions is that the number of receptor types is at least one greater than the number of degrees of freedom of the surface reflectance: there should not be more unknown variables than measurements (sensor types), so if the ambient lighting is one unknown, the human (trichromatic) visual system will be able to uniquely recover SSRs (i.e. exhibit perfect colour constancy) in situations with at most two degrees of freedom. This way, LMF enables specification

of the kinds of situations in which colour constancy is achieved by a particular perceptual system, and if it is not, what kind of non-constancy is occurrent. The range of cases in which constancy is occurrent for a given system can include greatly divergent sets of lighting, reflectance types and (perceived) colours. This range is said to comprise the collection of privileged environments of lighting conditions and surfaces for said visual system.

LMF was introduced in Maloney and Wandell (1986). Their justification for the assumption daylight has three degrees of freedom was derived from Judd, MacAdam and Wyszecki (1964). In this article, Judd and associates attempted to demonstrate that a satisfyingly accurate estimate of the spectral power distribution of daylight can be provided by the linear weighted sum of a limited number of basis functions (as few as three). The basis functions are orthogonal, meaning they represent independent degrees of freedom of the light; and the order in which the basis functions are derived corresponds to their proportionate contribution to the overall variability of the light.

From examining daylight measurements from various parts of the world, they found that the first most common basis function is the yellow-blue variation, corresponding to variations in cloudiness and ratio of direct sunlight. The second basis function expressed a pink-green variation, the variation corresponding to the water vapour content of the atmosphere. They determined the scalar multiples necessary to match these basis functions with a number of correlated colour temperatures, and found that three basis functions with the appropriate multiplicative factors offered a surprisingly good fit to spectral distributions that had been measured directly.

There have been many research projects to provide additional empirical justification of these claims regarding the structural components of daylight. These projects endeavour to ascertain what those structural components - if there are in fact any - might be. Wachtler, Lee and Sejnowski (2001) and Lee, Wachtler and Sejnowski (2002) describe such a project[NOTE 23], suggesting post-receptoral *opponent* processing is a highly efficient way to encode chromatic information, results that are compatible with the theses Shepard endorses.

Some additional physiological support for the claims made so far (about the 'filtering' of light in such a way that its structural components can, on some functional level, be distinguished) can be found in Van Hateren and Van der Schaaf (1998). In this paper it is suggested that if the function of simple cells (in macaque primary visual cortex) is to dissolve the linear superposition of signals that comprise an image into its independent component parts, certain properties of those cells should align with statistical properties of the environment[NOTE 24].

The central role of colour constancy in accounts such as the one described above has been criticised by several commentators, who claim that human colour constancy is decidedly poor (see e.g. Reeves, 1992). However, much of this counterevidence relies on tests involving test subject performance measurements with randomly chosen lighting conditions. Maloney therefore counters that finding results that indicate colour constancy for humans lies (significantly) below optimal levels in such experiments is not surprising: even if some visual system were to instantiate an LMF-type algorithm (and therefore exhibit perfect colour constancy across a particular range), randomly picking lighting and reflectance conditions to test that system's ability to achieve colour constancy is not likely to uncover the contours of its privileged environment, and will probably yield the conclusion that the system is not colour constant at all, or perhaps only approximately so in some cases. Compare an alien scientist using randomly picked wavelengths of electromagnetic radiation to determine whether a human test specimen can detect these. The band of visible light (for a human) is narrow, compared to the totality of possible wavelengths, so if the alien uses the entire electromagnetic spectrum to pick random samples from, the experiments are not likely to yield the information that for the 400 nm to 700 nm range, humans perform relatively well.

Shepard assumes that the LMF-model provides an appropriately close approximation of the process that yields colour constancy, and uses it to buttress his claims about the evolutionary basis for the properties of colour vision (Shepard, 1992a). He says the human visual system is able to recover reflectance and thereby achieve colour constancy (obviously within particular limits) because in evolution the trivariance of said system has adapted to the three degrees of freedom of terrestrial light. These three dimensions are light-dark (mid-day sunlight vs. moonless night or deep shade), red-green (light direct from the low sun, rich in long-wavelengths vs. said light-type as filtered by water-vapour-rich atmosphere) and yellow-blue (Rayleigh-scattering; light poor in short wavelengths in direct solar illumination vs. light rich in short wavelengths if object is blocked from direct sunlight but light scattered by clear sky falls on it). These three 'axes' are found to line up with those of perceptual colour space, which contains exactly the light-dark, red-green and blue-yellow opponencies (see section 3.2 for a more elaborate explanation of opponency).

All this provides some empirical support for Shepard's assertions, or at least for their tenor; the precision of fit of perceptual opponent axes and the degrees of freedom of daylight are a matter of some controversy. However, additional clues are to be found in the report of Delahunt and Brainard (2003) on their own experiments (which is in the same line of research as Judd et al. (1964)):

"The primary purpose of the experiments reported here was to assess whether the visual system’s adjustment to changes in illuminant (relative to a neutral illuminant) depends on the color direction of the illuminant change.

This question is of interest, since an analysis of the distribution of natural daylights indicates that some illuminant changes are much more likely to occur than others." (Delahunt and Brainard, 2003)

This is where we would need to locate the relevance of the claims Shepard makes: the components of typical daylight, the degrees of freedom of daylight corresponding to the three basis functions, are axes along which illuminant changes are most likely to occur. A visual system evolving to achieve the best results in such an environment would, in all likelihood, operate in such a way that its output space mirrors said three-dimensional structure. The suggestion that this is the default state does not preclude the possibility that some (or even many) species acquire additional representational dimensions[NOTE 25].

4.5.3 - The Ecological Hybrid Theory?

My suggestion is that Thompson and Shepard do not exclude eachother, but rather complement eachother quite nicely, yielding a hybrid theory of sorts. The way to elucidate this claim is by focusing on the *function* of colour vision. Both computational objectivists (such as Shepard) and Thompson speak of 'the function' of colour vision. Thompson says:

"I argue that the biological function of colour vision is not to detect surface reflectance, but to provide a set of perceptual categories that can apply to objects in a stable way in a variety of conditions. Comparative research indicates that both the perceptual categories and the distal stimuli will differ according to the animal and its visual ecology; therefore externalism and objectivism must be rejected." (Thompson1995b)

Shepard, on the contrary, would say that it is the function of the visual system to extract information about chromatic invariants from the distal scene. There is a deep-rooted ambiguity in the concept 'function' which feeds this conflict. However, I will argue that the two ways in which Shepard and Thompson understand and develop their notion of the function of colour vision can be complementary rather than mutually exclusive. I will maintain that these two accounts need eachother, that their dialectic might yield a better theory.

Wouters (2004) describes how Mayr (1961) engendered the opinion that *(evolutionary) biology* is to be distinguished from the other natural sciences based on its predilection and unique ability to ask and answer the 'why'-questions. On this view, one of those other natural sciences (i.e. the ones merely asking the 'how'-questions) is *functional biology*, described as a reductionist discipline borrowing heavily from physics and chemistry.

'How'-questions are answered by providing a mechanical description of some process, betraying a reductionist inclination amongst those who use this approach. 'Why'-questions are typically answered by reconstructing the chain of events of evolutionary history.

Pertaining to the function of colour vision, Thompson and Shepard fall on opposite sides of the ambiguity associated with the term 'function', and I claim that a great deal of the tension between these two evolutionary accounts of colour vision is due to this difference in interpretation. Thompson asks the biological/evolutionary 'why'-question, whereas Shepard asks the physical/evolutionary 'how'-question.

It does not make sense to ask the 'why'-question about lifeless physical objects (they do not acquire their characteristics through a selection history), and this is exactly Thompson's main point: as part of his Gibsonian inclination, he will claim it will not do to view a perceiving animal as nothing more than a physical object. However, there is nothing wrong with maintaining the physical perspective as *part* of a theory. Thompson should concede on this point, considering his naturalism: apart from the physical perspective deserving a part in a finalised ecological theory of colour, it makes sense to use the knowledge of physics as a methodological tool. In the end, the location of the label 'colour' might be a matter of convention, and in that decision-process, the pragmatics of the physical viewpoint might be a crucial factor, even if we shy away from scientific realism and merely adopt an instrumentalist perspective.

But now I am moving too fast. Returning to Thompson, it is possible to see that he reaches a similar (though weaker) conclusion at the end of his (1995b), when he redefines colour constancy (he suggests a *category* constancy rather than hue constancy), and claims that this way, the constancy phenomenon *does* play a major role in colour vision, and should be important to any theory of colour:

"I have argued that colour vision does not represent the world; rather, it presents the world to the animal by categorizing physically disparate stimuli into perceptual equivalence classes. Clearly such categorization is useful for the animal and so can reasonably be expected to play numerous further biological and ecological roles. (...) In any case, whether it is due to natural selection and/or other types of evolutionary factors, colour constancy in the sense discussed here figures largely in human colour vision and probably does so in the visual ecology of other colour-seeing animals."

Note the use of the term 'biological role' – this is not a neutral composite term, as Wouters' (2004) definition demonstrates:

"The biological role of an item or activity is the way in which it contributes to an activity or capacity of a larger system." (Wouters 2004)

Unpacking this concept, he specifies an overview of the kinds of questions habitually asked in organismal biology. He says that an item or behavioural pattern's function is defined in terms of...

"(1) the form and activity of that item or behavior (description)
(2) its biological roles,
(3) the causes and underlying mechanisms resulting in the performance of those roles (Tinbergen's 'causation'),
(4) the biological value of that item or behavior having the character it has and of the performance of that role (Tinbergen's 'survival value'),
(5) the development of that item or behavior in the course of the ontogeny (Tinbergen's 'ontogeny'),
(6) the origin and modification of that item or behavior in the course of the evolution (Tinbergen's 'evolution')." (Wouters 2004)

My suggestion is that for colour, answers to these question-domains will look like this:

(1*) a description of the visual system in physiological and neurological terms;
(2*) a behavioural explanation: how does the organism use colour vision?;
(3*) by what mechanism does the visual system perform the function belonging to its role as specified under (2*)? There are two kinds: explanations in terms of cause and in terms of mechanism.
    (3*a) causal: reconstructing the causal chain of stimuli and responses;
    (3*b) mechanistic: describing the functioning of the whole in terms of its parts;
(4*) specification of the survival utility of colour vision:
    (4*a) why is it useful for an organism to possess colour vision?
    (4*b) additional question: why is it useful to have that particular kind of colour vision - say, tetrachromacy as opposed to trichromacy?;
(5*) a description of the ontogenetic development of the visual system;
(6*) a description of the phylogenetic evolution of the visual system.

The Linear Models Framework provides a functional answer to (1*). Shepard does this by first suggesting the hypothesis that LMF is capable of providing an answer to (3*a) (e.g. describing the mechanism involved in colour constancy), and based on this answer he tries to find answers to (4*b) (an organism has become attuned to large-scale invariants in the optic array; for humans, this would help explain their trichromacy) and perhaps the first steps towards an answer to (6*) (by pointing out ecological contraints to the evolution of the visual system, in terms of the degrees of freedom of terrestrial light).

Thompson, rather, starts out at (2*), by wondering what the ecological function of colour vision might be for a wide variety of species, and he answers that question by stating an animal should exhibit behaviourally relevant colour *category* constancy. The Comparative Hypothesis can then be a natural response to the great variety of answers to (3*b) in different animals (even if only because of divergent retinal dimensionality across humans), which enables Thompson to formulate a response to the more

general (4*a), for instance by referring to the work of Mollon (see section 3.6.3), in a sense coming full circle to (2*).

It is clear Shepard and Thompson are asking and answering different (sub-) questions concerning the function and functioning of colour vision. Therefore, it is not surprising their accounts appear to diverge on a number of essential issues. However, it is quite interesting to see how together they cover almost the entire spectrum of possible questions about the function of colour vision: as explained above, Shepard's theory touches upon (1*), (3*a), (4*b) and (6*), whereas Thompson's account concerns (2*), (3*b) and (4*a).

If we take a look at the different meanings of 'function', we see a similar division of labour as the one highlighted above. Wouters (2003) distinguishes four kinds of function:

"(1) function as activity ($function_1$), (2) function as biological role ($function_2$), (3) function as biological advantage ($function_3$), and (4) function as selected effect ($function_4$). $Function_1$ (activity) refers to what an item does by itself; $function_2$ (biological role) refers to the contribution of an item or activity to a complex activity or capacity of an organism; $function_3$ (biological advantage) refers to the value for the organism of an item having a certain character rather than another; $function_4$ (function as selected effect) refers to the way in which a trait acquired and maintained its current share in the population."

Applying these distinctions to the colour case, we can see that Shepard defines (4) in terms of his answer to (1): (human) trichromacy was selected for in evolution because it offered sufficiently reliable colour constancy in the face of the specific structure of environmental illumination. Thompson focuses on (2) by posing (3) as a question with a deeply relativistic answer: the fact that different animals in different niches can possess different kinds of colour vision has certain implications for the role that colour can be said to play in an animal's interaction with its environment.

The above shows that the range of the kinds of questions of one complements the range of the other's investigation into the subject matter in a rather interesting fashion. The first thing we still need to do is to demonstrate how Shepard's functionalist line of reasoning could be relevant to a question of *ecological role* - if this can be done, the two approaches could possibly be seen to coagulate into a useful hybrid theory.

Once more taking a look at Wouters (2004), he argues that functional biology, with its descriptions of the workings of organs and other biological components, is not un-biological in the way evolutionary biologists might claim it is (which would be because of its reductionism). Rather, the functional investigative path allows us to understand the contributions of the various organs to the organism's life state - that is, the functioning of the parts in the context of their contributions to the continued existence of the organism as a whole.

If we look at the four interpretations of the notion ‘function’ again and parse them in terms relevant to the issue of ecological colour, Shepard’s answer to (1), understood as an explanation of the contributions of the animal's various subsystems (the chromatic aspect of the visual system, in this case) to the organism's life state, might yield the basis of an answer to (2). In other words, if we use a functionalistic answer to the question ‘how does colour vision work?’, we might construct the theoretical and methodological framework with which to look for an answer to the question ‘why does organism X have a colour vision system of type Y?’.

To understand how this entanglement of the accounts of Shepard and Thompson might come about, we need to return to the quote used at the beginning of this section:

"I argue that the biological function of colour vision is not to detect surface reflectance, but to provide a set of perceptual categories that can apply to objects in a stable way in a variety of conditions. Comparative research indicates that both the perceptual categories and the distal stimuli will differ according to the animal and its visual ecology; therefore externalism and objectivism must be rejected." (Thompson1995b)

Thompson advocates a kind of colour category constancy here. Now, this acceptance of constancy is in itself already an interesting cross-pollination with what computational objectivists are attempting to do. But, there is more.

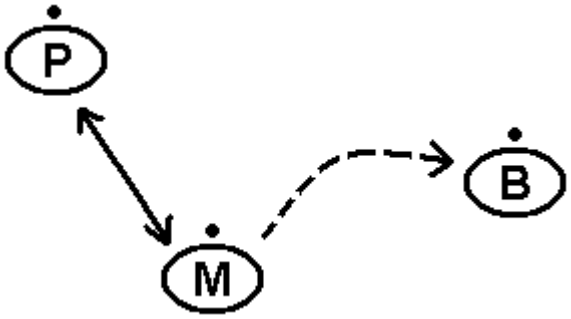
Thompson’s Gibsonian inclination might have serious consequences for what he means by ‘stimuli’ in the above quote. If Thompson understands the stimulus to be detected as an affordance (which his insistence on colour categories as modes of presentation rather than the computational objectivist’s representation (1995b) does suggest), this would have a serious consequence. Such a choice by Thompson will fail to help him discredit the possibility that the actual mechanism that yields the chromatic experience utilises SSR (or the related notion 'productance', as suggested by Byrne and Hilbert, 2005) and colour constancy. This point once more rests on the ambiguity of the notion ‘function’, and how Thompson asks a very specific and one-sided question, for which Shepard provides the (equally one-sided!) counterpart. Therefore, it might be the case a story like Thompson's actually *needs* a story like Shepard's for a complete story about colour vision, and vice versa.

4.6 - Towards Colour Concepts

I have taken quite a bit of time to make a very important point. I wanted to show how properties of the agent and properties of the environment are fundamentally entangled - *co-attuned* - to collectively yield colour-discriminating behaviour. Objective properties of the environment (e.g. the structure of terrestrial illumination) and objective properties of the agent

(e.g. his retinal dimensionality), together with the way these properties relate to eachother via the agent's actions, are all essential to a complex interaction dynamic of affordances and effectivities.

Expanding upon the classification into various domains hinted at in section 3.5, one way to depict this interrelatedness of agent and world is in the following, simple diagram:



[Figure 6: interrelatedness of agent and *physical* world]

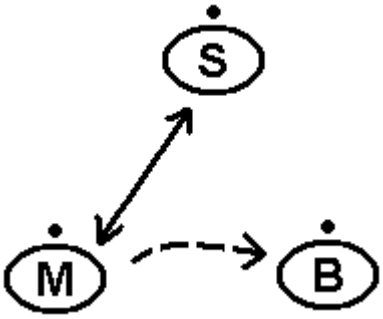
That is, the biomechanical properties of the agent as they change over time (dM/dt), in *interaction* with physical properties of the environment as they change over time (dP/dt), can, at some level of detail, be *described* as behaviour (i.e. the change of behavioural patterns over time, dB/dt). This behavioural description can be expressed in terms of a behavioural space, as a higher-dimensional version of the movement planning field utilised by Thelen et al. (2001). The discussion about ecological colour perception can be understood as an attempt to provide a qualitative specification of how these agent-properties and physical environment-properties hang together, at least in the case of colour perception.

However, certain rather important aspects are missing from the ecological story about colour perception. One essential factor that does not appear to fit into the ecological story is the *socio-cultural structure* we are immersed in. This is odd, because from an enactivist perspective, or even from a broadly $E_{(i)}C$ viewpoint, it would appear prudent to include such things as *language* and *social structure* in the suite of meaningful scaffolds which (at least human) agents help co-constitute, and by which they are influenced in a profound manner. In other words, my claim is that a theory of $E_{(i)}C$ cannot afford to remain silent about the affordance-effectivity[NOTE 26] push-pull-system involving socio-cultural practices of which each human agent is a part, and which does not, at first glance, appear to be fundamentally different from the affordance-effectivity push-pull-system involving agentive and environmental properties (at least in structural terms). The discussion about socio-linguistic influences on colour phenomenology (in section 4.3) provides clues to solve the other half of the puzzle.

That is, the ecologically inclined theories of Thompson and Shepard on colour perception contribute to an $E_{(i)}C$-appropriate characterisation of

agent-environment interaction (at least to the extent that this interaction involves colour, but I assume the general case is appropriately analogous). The idea in section 4.5 was to demonstrate how properties of the animal are attuned to properties of the external world, hence how both these sets of properties *collectively* yield the animal's colour-perception-based interaction dynamic with its surroundings. This is exactly the point of figure 6, depicted above.

On the other hand, the ideas to be derived from the discussion about colour language and culture involve the co-attunement of agent and socio-cultural environment. Depicted in a diagram like the one above, this looks as follows:



[Figure 7: interrelatedness of agent and *social* world]

That is, the biomechanical properties of the agent as they change over time (dM/dt), in *interaction* with socio-culturally determined properties of the environment as they change over time (dS/dt), can, at some level of detail, be *described* as behaviour (i.e. the change of behavioural patterns over time, dB/dt). This behavioural description can be expressed in terms of a behavioural space, as a higher-dimensional version of the movement planning field utilised by Thelen et al. (2001). Section 4.3 can be understood as an attempt to provide a qualitative specification of how these agent-properties and *social* environment-properties hang together, at least in the case of colour perception. Chapter five below will contain a more elaborate exploration of the factors depicted in this schema.

There is a very important point to be made about the idea that the Thompson/Shepard-hybrid and the theories involving the linguistic anthropology of colour are to fit together: Thompson's Gibsonian inclination in particular suggests a very specific role for the notion that phenomenal colour space has a structure. This structure should be understood *not* as a representational structure of the cognitivist kind (i.e. an internal representation of external events), but as an expression of *behavioural* responses ('B' in the schemata above) to environmental prompts ('P'), based on a particular physiological tendency ('M'; e.g. the ways in which the visual system works[NOTE 27]) plus sociocultural constraints ('S'; e.g. the fact that linguistic rules or social customs force behaviour in a particular direction). The status of 'representation' in $E_{(i)}C$-based descriptions of this

kind will be discussed in chapter 7; for now, the main idea to take away from this discussion is that the integration of talk of 'phenomenal structure' into the theory of concepts I am currently building need not imply cognitivism.

The basic idea developed in the past chapter aligns, in some sense, with George Lakoff's description of what he understands 'colour' to be:

"Color categories result from the world plus human biology plus a cognitive mechanism that has some characteristics of fuzzy set theory plus a culture specific choice of which basic color categories there are" (Lakoff 1987)

Apart from certain aspects of this claim that I take issue with ('cognitive mechanism' and 'fuzzy set theory' in particular - see chapter 7 about the former, sections 6.11.2 about the latter point), the main problem with the 'explanation' in this quote is its profound opacity: it sounds great, but what does it *mean*? The comparison of the ideas of Thompson and Shepard above provided a few clues about the 'the world plus human biology' bit, the discussion about culture-centric colour language involved topics is in line with Lakoff's 'cognitive mechanism' and 'culture specific choice'.

There is still some work to be done to determine what these ideas, leading to the subsequent discussion of what this might signify for a *cognition*-based agent-environment interaction dynamic (rather than 'merely' perception-based, as in the foregoing). The next chapter will be devoted to providing a more detailed description of the structure and development of colour space, in a way that is useful to the subsequent expansion of this idea into a theory about concepts.

= =

[SUMMARY of chapter 4 AND PREVIEW]

The standard account of colour perception involves a phenomenal structure that is defined in three dimensions (brightness, saturation and hue), as derived from behavioural responses to environmental prompts, and informed by the properties of the human retina and the subsequent neural processing of chromatic stimuli. This behavioural manifestation includes a cultural and linguistic aspect: different languages do not contain the same number of basic colour terms, despite the universality of the neural substrate. Critics of this received view, often supporters of linguistic relativism, tend to draw attention to the fact that this view is guilty of severe *decontextualisation*: in several 'primitive' cultures, the basic colour words which express basic colour categories are not neutral labels, but rather include many other semantic aspects. That is, there are sociocultural environmental influences in play as well.

Based on these considerations, I defend an intermediate position which underlines the importance of *both* a pan-species neurophysiologically-

based tendency towards certain categorizations *and* the importance of sociocultural influences. This view suggests that there is a connection between basic sensorimotor contingencies and higher-order sociocultural regularities. That is: that basic three-dimensional phenomenal structure which describes the behavioural responses as informed by neurophysiological properties should be expanded to include more complex behavioural responses as informed by sociocultural properties.

If these ideas are combined with the dynamical movement planning field as a description of basic enactive behaviour from chapter 3, this yields the first notion of a more complex behavioural space, which in the chapters to come will be described more extensively, and transformed into a *conceptual* space.

But first: there is a third influential realm of properties, apart from the neurophysiological (body-based) and sociocultural (social environment-based) properties mentioned above: physical environmental properties. In this past chapter, we have seen that a promising enactive account of colour perception (the one by Evan Thompson) can be augmented with an ecological theory of colour perception which focuses on the idea that there is a congruence of a perceiving agent's retinal properties and environmental regularities, namely the chromatic structure of the optic array (Roger Shepard's view). This idea suggests that there is a mutual attunement - of agent and environment - which helps determine the agent's perceptual and behavioural possibilities. This agent-physical environment-interaction can be combined with the agent-social environment interaction (both of which depend to an important extent on the agent's embodied properties), to provide a fairly well-rounded description of the structure of the agent's $E_{(i)}$C-related behaviour.

That is: the enactive/situatedness-notion itself suggests the relevance of social (in addition to ecological/environmental) affordances in explaining the behaviour of an agent; the universalist linguistic account of 'colour-cognition' runs the risk of decontextualisation, so a moderate relativism (the involvement of sociocultural factors to ameliorate the body-based properties of the agent) is needed. Both of these approaches combined cover the entirety of the agent-environment interaction dynamic. In chapter 5, the properties of the basic colour space will be explored further, in preparation of the introduction of conceptual space in chapter 6.

= =

[5 - The Structure of Concepts]

5.1 - Progressive Segmentation Of Colour Space

In the coming sections, I will describe ideas by Kimberly Jameson about the structure of phenomenal colour space, and the ways in which this structure can develop to accomodate changes in a person's linguistic apprehension of colours. These ideas will help characterise the interaction of an agent with his *socially* constructed environment - the second component (after the interrelatedness of agent and *physical* world, depicted in figure 6, section 4.6) of the full $E_{(i)}C$-appropriate story about an agent's interaction with his environment. After this, in chapter 6 and on, I will use these ideas to develop a more general story about the properties of conceptual space.

In an earlier section (4.3), I offered a description of the ideas of Berlin and Kay (1969) concerning the linguistic categorization of perceptual colour space. Their claim was that the influence of language and culture on colour perception was minimal: the locations of the hue foci in perceptual space are determined by physiology, and are roughly the same for all 'normal' humans. In essence, they can be said to claim that the Sapir-Whorf-thesis (which predicts the transformation of thought by linguistic structure) is incorrect.

Opponents of this view (e.g. Saunders and Van Brakel, 1997) tend to accept the Sapir-Whorf-thesis, and one of the important points entered into the discussion on the basis of this adherence is that in many primitive languages, colour words are not user- and context-neutral in the way that English colour terms are.

Recall (from section 4.3) that for the Hanunóo, and many other linguistic communities, words denoting colour-relevant information are necessarily linked to other properties - Hanunóo not only encode colorimetric information such as oppositions between light and dark, but also between dry and wet, or deep/unfading/desirable and pale/colourless/weak. If we were to depict this in a perceptual colour space, this would mean the expansion of that three-dimensional model with additional dimensions (in order to depict the parameters that specify the properties linked to such multireferential colour terms), to capture the semantic linkages and content-ascriptions these people make[NOTE 28].

This expansion of perceptual space with non-perceptual (or not-necessarily-perceptual) properties can serve as a first step towards constructing a *conceptual space*: a topological, multidimensional space that depicts the concepts an agent possesses[NOTE 29]. This conceptual space will form an important component of my theory about concepts. I intend this theory to be $E_{(i)}C$-compatible, and because (in classical cognitivist theories) concepts are considered to be the building blocks of thought (that is, including, or perhaps *especially*, 'higher' cognition), I submit my theory will also be, in some sense, a theory about higher cognition in an $E_{(i)}C$ framework.

Towards that end, this chapter will be dedicated to finding a proper theory of perceptual (and conceptual) space categorization. I will base such an account, in part, on Kimberly Jameson's 'Interpoint Distance Model-Framework'.

Jameson (2005) proposes to use, defined in order of prominence, a compound lightness - saturation - hue criterion as a basis of colour space segmentation, and she shows how a colour space segmentation mechanism can produce universal linguistic categorization tendencies despite differing psychophysical dimensionalities (for instance, di- or tetrachromatic humans).

5.2 - The Interpoint Distance Model

In her 2003 manuscript "Culture and Cognition: What is Universal about the Representation of Color Experience?", Kimberly Jameson describes a more evolved account of the ideas presented in her 1997 article together with Roy D'Andrade (see section 4.3 for a discussion).

Her Interpoint Distance Model (IDM) framework involves a suite of mechanisms she claims to be involved in the cultural-linguistic segmentation of perceptual colour space. Her goal is to clarify the complex entanglement of the colour-related aspects of culture, cognition, language and neural processing. As an interesting side note, one that cements the relevance of this model to my own account, my goal in this book is the same, only in a broader context: I wish to clarify the complex entanglement of culture, cognition, language and biomechanical processes, within an $E_{(i)}C$ context.

Jameson's account adheres to three premises:

(1) the segmentation process is not fueled or determined by just hue salience;

(2) regarding the determinants of phenomenal colour space, hue is not even the primary factor; instead, the lexical encoding occurs according to the ordered sequence [a] brightness, [b] saturation [c] hue;

(3) cultural, linguistic and additional environmental factors may exert influence on the colour space segmentation evolution.

The first thing to note about Jameson's approach is that she appears to start out at the default universalist position, stressing the importance of a particular organisation of perceptual space (per point (2) above, she claims hue is not the primary aspect, but there *is* a particular structure in place), and then adds a relationalist spin.

The second interesting aspect of IDM is the kind of argument provided in favour of point (3). This argumentation kicks off with the claim that the influence of variations in retinal sensory dimensionality (see section 4.1) should not be underestimated. Apart from the familiar brands of dichromacy (yielding what we call 'colour-blindness'), and aberrant versions of trichromacy (where the sensitivity spectra of one or more of the cones might stray from the norm), Jameson claims as much as 20% of Caucasian females might be retinal *tetra*chromats, or exhibit the genetic potential towards such increased receptor-dimensionality. This would mean that the people in which said potential is actually realised might be able to discern more colours than average trichromats, although this would be contingent upon the specifics of the post-retinal processing in these individuals.

Now, if these variations in perceptual dimensionality can occur *intra*-culturally, the question regarding the significant *inter*-cultural agreement on the linguistic segmentation of colour space (as defended by the universalists in the tradition of Berlin and Kay) presents itself once more with unprecedented force.

Jameson theorises that the structure of our colour *language* exerts a converging force upon the relation between colour perceptions and colour terms in these anomalous colour perceivers. Despite the fact tetrachromats might be able to discern certain colour shades, they do not possess the words to label these novel perceptions. That is, because they are forced to live in a world geared towards trichromatic perceivers, with as a determining feature a trichromacy-based colour language, they have devised cognitive procedures to map their higher-dimensional colour perceptual space onto the coarser-grained linguistic structure of the trichromats. A similar strategy might be employed by other anomalous perceivers (such as dichromats), who would need to project their differently-structured perceptual space onto trichromatic linguistic space, with the use of an intermediate cognitive layer.

The point to take note of here, which embodies the explanatory shift Jameson suggests is the component of her model that does the work, is the distinction between, on the one hand, the formation of a specific culture's colour lexicon, and, on the other hand, the way in which individuals within that culture use that lexicon. The latter aspect would then include the perception-to-lexicon-transformations already mentioned, as an expression of the dynamic linkage between these perceptual and lexical levels.

Let us now take a closer look at the way Jameson uses the IDM framework to explain the processes of progressive segmentation of perceptual colour space. On pages 29-31 of her article, she proceeds to list a number of principles which her IDM framework proposes. Principle (1) is:

"1. The cognitive dimensions (ordered by importance) Lightness, Saturation, and Hue are primitives in both individual and cultural color representations. However, Lightness and Saturation are of paramount importance in the initial stages of a culture's color naming system." (Jameson 2005)

Jameson claims that the empirical data provided in support is compelling: at the very least, brightness is a factor of great importance in the early stages of evolutionary colour space categorization. Arguments in support of saturation as an independent contributing factor are less powerful: it turns out to be the case many test subjects find it difficult to uncouple brightness from saturation.

Principle (2):

"2. Cultural color naming systems and categories develop through successive partitioning of an idealized normative color appearance space on the basis of the dimensions given in (1). Category partitions in such cultural systems strive to satisfy two equally important goals: optimization of polar symmetry and category-area uniformity and balance relative to the cognitive dimensions in (1), and responsiveness to socio-cultural-environmental pressures such as demands for representational specificity of color, demands for a non-idiosyncratic (or normative) color information code, and compatibility with existing ethno-linguistic structures. The implementation of principles (1) and (2) results in a color naming system that is effective for the communication needs of the users of the system." (Jameson 2005)

This is the notion that surfaced in the 1997 article Jameson co-authored with Roy D'Andrade (see section 4.3). The basic idea is that every subsequent segmentation iteration within that three-dimensional perceptual colour space should result in a new colour name that is most informative to the speaker in his particular environment and socio-cultural embeddedness. As a general rule, the colour name (identifying some brightness / saturation / hue-focus) that is furthest away from the colour foci already named will possess the greatest informative value; this procedure will yield something resembling the familiar evolutionary colour name sequence (Berlin and Kay 1969, Kay and McDaniel 1978). However, external constraints might favour the expedited naming of a different colour focus, e.g. a prevalence of green over yellow in a forest environment might cause the green focus to be lexicalised before the yellow focus.

Jameson provides some additional details concerning this colour space segmentation strategy in principle (3):

"3. Individual color naming systems and categories first arise through learning a culturally normative naming system and its relation to one's individual (personal) perceptual color appearance representation. The individual's perceptual color representation is related to the culture's color naming system through a color naming-function (Alvarado and Jameson 2002). Over an individual's lifespan a personal naming-function evolves (e.g., new category labels are learned), which relates the culture's normative naming system to the individual's perceptual representation." (Jameson 2005)

This principle contains an essential point, which I would like to emphasize a little more than Jameson does to underline how IDM stays away from orthodox cultural relativism. Here, I propose we listen to Don Dedrick (1997), who stresses the importance of the perceptual saliencies of hue foci as a pregiven structure for the cognitive / linguistic schema to latch on to. These saliencies are, to an important extent, *physiologically* determined.

There is room for idiosyncratic variation on this account: true human (full-blown perceptual, not just retinal) tetrachromats[NOTE 30], if they exist, will obviously diverge from the norm, as will dichromats. Also, there is no objection to slight variations of the exact location in colour space of hue foci from one regular trichromat to the next, whatever the cause of this could be (to the extent that it depends on the person's physiology). The point is that, for every individual, there are facts of the matter about the role his physiological makeup plays in which colour shades are considered good exemplars of a specific category, and which are not.

This will not cause IDM to devolve into the familiar universalist account, for there is still a highly significant role to play for non-physiologically determined factors: language, culture and idiosyncratic preferences can also exert their influence. This means that a person's cognitive processes work in tandem with the performance characteristics determined by the individual's physiology.

An important addition to the IDM framework that I will develop (in particular in chapter 6) involves an extrapolation of perceptual space with *conceptual* space, and this expansion should offer more room for the inclusion of socio-cultural, linguistic (and so on) factors, and a more substantial account of the differentiation of perceptual space-based colour names according to the various environmental constraints a speaker is subject to. An important *modification* of mine of Jameson's model will be the suggestion of a model or mechanism that achieves the same explanatory goals, but *without* the need for explanations in terms of ubiquitous and overt cognitive processing - i.e. an account of conceptual space-dynamics that is congruent with $E_{(i)}C$.

But for now, I shall continue with a discussion of Jameson's ideas. She claims the three principles discussed above yield a number of important consequences. Consequence (1) states:

"1. Because lexical categories are progressively assigned in ways that tend to maximize information content and minimize label-to-exemplar confusability in communications between members of a culture, the naming system developed will necessarily depend on the range of colors available, extent of each color represented, and the ordering properties (discrete or continuous) of the stimulus space to be named. These features may differ, as when belonging to two natural environments (tropical versus desert), and will differ between two scientific color-order systems (Munsell versus CIE). IDM theory suggests that a space with a non-regular distribution of items

across categories (e.g., an unusual space with a large yellow stimulus region, compared to a much smaller red region) will be named in a manner that accounts for the color region 'bumps' in the space (Jameson & D'Andrade 1997) regardless of whether it is a manufactured stimulus space or a natural environmental color space." (Jameson 2005)

These remarks serve to define the colour space to be segmented. There is a hint of environment-dependent differentiation between different flavours of colour space here: there would be room, within the IDM framework, to incorporate the perceptual colour spaces of people with physiological adaptations to various environments, with its own spectral arrays: consider, for instance, the adaptation of yellow-tinted lenses found amongst some humans living in areas of the world with brighter-than-average sunlight. Their perceptual colour space would certainly differ from the norm, as it would have the colour region bumps and indentations referred to in the quote above at places where people from less sunny locales would not. In particular, such people would be less sensitive to differences between blue and green (Hardin 1988/1993, pg. 167). Note that 'normal' perceptual colour space already has an irregular space in most universalist accounts.

Therefore, this kind of variation is likely to be limited enough to fall within the range of 'allowed' cases for a *quasi-universalist* account, where it can still be maintained that there is, for every individual, a fact of the matter about how his colour perception system processes incoming stimuli, resulting in a particular colour saliency structure (which I dub the 'neurophysiological yield')[NOTE 31]. Considering the likelihood that the variance in colour spaces across individuals worldwide is expected to be fairly narrow (a largely convergent genetic base will result in overwhelming percentages of trichromatic subjects), combined with the convergent forces of a shared trichromacy-based linguistic encoding scheme (the prominence of which Jameson herself stresses - see above), the acceptance of this slight possible differentiation in colour space flavours does not in and of itself do any major damage to the universalist's theory.

However, consequence (2) presents a more powerful case for the inclusion of relativist aspects into a proper account of colour space segmentation:

"2. Category regions, and interim category best exemplars, change as a culture's color naming system develops and successively defines new category partitions. Category focals thus shift and as a result are salient only as a function of the unfolding of the partitioning process." (Jameson 2005)

This suggestion is closely akin to Bernard Harrison's point about the embeddedness of colour naming practices in a wider linguistic relational web[NOTE 32]. Now, according to the modification of IDM that I wish to develop, focal shifts are caused by cultural-linguistic transformations, but the allowed deviation of each of these is constrained by the characteristics of the underlying physiological system. It does not seem likely that, for example,

the force of linguistic change in a given culture with a constant majority of regular trichromats would be such that the green focus would venture into the domain of perceptual red. The advent of overwhelming numbers of functional tetrachromats (or dichromats, or anything other than plain trichromats), to such an extent that the majority of the population would come to consist of these mutations, might change this situation. In this case, most people would no longer consider the familiar colour language to be sufficiently correct, and the linguistic schema would be modified to reflect these physiological changes. However, such a shift does not appear to be on the verge of occurring.

"3. Because the constraints of principles (1) and (2) above are universal across cultures, the evolution of color naming systems will converge somewhat, producing general features of color naming that are universal across cultures." (Jameson 2005)

This sounds plausible, and aligns with what I have been claiming. However, it might be the case that the (relative) universality of colour categorization as demonstrated by vast amounts of empirical data will not be sufficiently secure based on Jameson's principles (1) and (2). She describes these principles as (mostly) *cognitive* strategies, and while similarities in human behaviour (and the universal constraints encountered in the environment) independent of cultural descent will undoubtably contribute to the similarities in colour lexicalisation across cultures, the emphasis placed on intra-cultural differences (for instance, in terms of environment, or colour ordering system used) under consequence (1) suggests something else is needed. An alternative suggestion is to play up the universalist aspect of the theory, i.e. by stating that - for humans, obviously - physiological similarities can close much of the gap, and help explain the (relative) pan-species universality of colour categorization. The vast majority of the world's population is trichromatic, exhibits behaviour consistent with opponency theory (i.e. if tested with the Hurvich/D. Jameson method; see section 4.1), and so on, and these facts should play a large role in explaining the convergence of empirical data gathered across the world.

An important question, relevant to this particular issue (i.e. of a mainly physiology-based convergence of colour spaces), which arises, is: is tetrachromacy really as widespread as Jameson suggests? The litmus test would be whether these supposed tetrachromats are able to watch television and see 'normal' colours. No amount of cognitive processing aimed towards linguistic convergence (e.g. the four-to-three-dimensional mapping Jameson suggests for tetrachromats in a trichromatic world) will help a tetrachromat see what we see when we watch television.

Still, Jameson wishes to maintain this cross-dimensional mapping will suffice to smooth over any practical (i.e. communication-based) differences between humans of various retinal dimensionalities (at the very least for dichromats in a trichromatic world), as per consequences (4):

"4. Even though color appearance representations for individuals from the same culture may differ, the individuals can share and effectively use a normative color naming system." (Jameson 2005)

... and (5):

"5. Individual color naming can reflect differences in personal color appearance representations (e.g., different category foci can be found across individuals; see the collected work of MacLaury), yet social practices of 'linguistic charity' (Putnam 1988) permit some variability in individual color naming and perhaps expect it from the probabilistic features of the gradient stimulus space (c.f., Kay & MacDaniel 1978)." (Jameson 2003)

These two consequences express the normalising effect of partaking in a shared socio-linguistic community: ideas pertaining to the appropriate application of a word will, across speakers, tend to converge over time, simply because if a word is used in an unconventional manner, the speaker will fail to get the response he expected or desired. This is the extent to which I would wish to support Jameson's notion of cross-dimensional-mapping: as it is, as yet, rather unlikely that there are many operational human tetrachromats, for now the apparent success of these strategies of cognitive convergence for perceptual colour space across cultures and linguistic communities is based, for the most important part, on the similarities in neurophysiological yield. Hence Jameson's cross-dimensional mapping should be understood as a general form of the more specific same-dimensional, language-to-language mapping (for each individual ranging over a shared similarly-structured perceptual space).

Differences in colour-related physiology between persons, compounded by idiosyncratic naming practices will, for the most part, be smoothed over by the linguistic convergence mechanism described under consequence (4). Any differences still left after that will probably fail to result in an opulence of practical problems: obviously, tasks in which the best example of a particular colour are to be picked out do not arise very often in everyday life. Usually, general agreement about the lexical coding appropriate to a specific colour is sufficient: the lack of consensus pertaining to the exact degree of similarity to a hue focus is vastly less important than the fact we all know to use colour word X when confronted with perceptual stimulus falling somewhere within (fuzzily bounded) range Y. In many potentially problematic cases, the words used (i.e. greenish blue, salmon-coloured) are unspecific enough to veil any occurent idiosyncratic perceptual variation.

However, the mechanism of cross-dimensional mapping Jameson suggests will become more relevant in the case of conceptual space; this notion is to return in chapter 6.

Jameson's Consequence (6):

"6. Although the cultural development of a color naming system evolves category partitions by following the principles stated in (2) above, it may undergo successive re-partitioning in response to social pressures. (For example, a need to now differentiate blue and green separately from a previously defined GRUE category)." (Jameson 2005)

One of the anti-universalist points of criticism provided by Saunders (1992), was that colour names in some cultures tend to be linked, semantically, with specific objects, organisms, substances or cultural practices. I made a similar point in section 4.3, based on suggestions by John Lucy. It is conceivable that Jameson's account would leave some room to deal with such cases, up to a point (as per Jameson's consequences (2), (4) and (5), and (6) immediately above). Hence, Saunders' more specific criticism that westernised abstract colour words are inappropriate as glosses for many colour terms used in more 'primitive' cultures is much less compelling when directed against Jameson's account than it is against orthodox universalism: IDM explicitly incorporates the possibility (and even probability) that the referential range of colour terms differs between individuals. All this does not change the important physiological similarities (in terms of their colour perception system) across humans everywhere.

However, both in Jameson's conception of IDM and in the emphasis-shift currently defended, the perceptually salient hues are not the be-all end-all of colour lexicalisation - the cognitive and conceptual levels incorporate cultural-linguistic influences. This means that perhaps consequence (6) of IDM described above could offer a way to incorporate the linkages Saunders highlights, since IDM acknowledges that influences from outside the strict confines of the colour domain as such can be highly significant.

So Jameson can incorporate the contributions of environmental pressures in what constitutes maximised saliency in the evolutionary order of colour space segmentation: e.g. a choice to encode either yellow or green after red might not depend solely on the choice between chromatic or lightness information increase (or saturation, or optimal category size), but perhaps on environmental properties (ubiquity of either yellow or green foliage, making one or the other a more practical choice to have a word for). Hence, a specific colour that is overwhelmingly prevalent in some linguistic group's environment, and/or linked to a highly significant object or custom in the group's culture, might be classified in its own colour category much sooner than the evolutionary stage of the language in question would suggest.

The final entry in Jameson's list of consequences, (7):

"7. When differentiation on the basis of lightness and saturation has been optimized, successive repartitioning will proceed using principles in (2) and Hue in novel category formation." (Jameson 2005)

This is, in essence, what MacLaury (1997) claims: for lower-stage cultures, brightness might be the dominant categorization paradigm, but as both society itself and the partitioning of colour space used grow more complex, using hue categories emerges as a more efficient strategy.

This might also imply an explanation for the error Berlin and Kay (1969) have been chastised for making: for colour space segmentations in Western languages, hue is indeed the primary determinant for colour foci. However, if we follow Jameson and MacLaury on this point, utilising this criterion to backwards-engineer the evolutionary order for the linguistic segmentation of perceptual colour space, one is in danger of misrepresenting, to some extent, the particulars of colour perception and naming in the more 'primitive' cultures, who do not utilise hue the way speakers of (say) English, or Dutch or German do.

The IDM framework can account for much of the data, and offers sufficient explanatory 'heft'. However, from the discussion of the various principles and consequences above, I feel it is obvious I need to make some modifications to make this account work.

The first step is to attach greater importance to the relative universality of neurophysiology of human colour system. The exception-cases Jameson mentions are not necessarily a factor of overwhelming influence. For instance, the percentage of tetrachromats is likely to be much smaller than her claim of up to 20% - the television-counter-argument mentioned under the discussion of consequence (3) above would have to be met in a convincing manner to substantiate such high estimates.

Another problem for Jameson's account is the problem of covariance, often even practical indistinguishability of the lightness and saturation dimensions. Recall that Jameson emphasizes the important causal role of both the brightness and saturation dimensions in colour categorization. A problem with this idea is that it is notoriously difficult to separate these two factors in regular perception tasks, something Jameson (2005) herself acknowledges. However, Jameson argues that certain phenomena, for example the occurrence of the ‘GRUE'-category, are more easily explained if variations in lightness rather than hue are utilised as providing a classification impetus.The lumping together of green and blue in a single category by speakers in early stages of the evolutionary colour categorization sequence can be explained if, for those people, the similarities of lightness and saturation of these two colours are thought to be more significant than the difference in hue. Yellow, for instance, can be easily distinguished from blue and green in terms of lightness and saturation, and would, if the lightness and saturation criterion were to be used, be a likelier candidate to be placed in its own category in an early stage.

This is a fairly compelling line of reasoning, and parallels arguments offered by MacLaury (1997), who suggests the brightness (what Jameson calls

'lightness') and hue sequences will merge at some point in the evolutionary development of a language's colour vernacular. However, this argument does not need the potentially problematic separation of causal contributions of the brightness and saturation dimensions, and works quite well if only brightness is used.

This concession does not, on its own, imply that humans from 'primitive' cultures perceive colour differently (i.e. that their neurophysiological yield would be different), merely that in their physical and socio-cultural environment, colour space segmentation based on lightness results in more informative colour names. The salience of classifications along the hue dimension appears greater to cultures with a more sophisticated or complex colour vernacular (in terms of the position within the evolutionary sequence), and overseeing the *totality* of colour space it appears that differences in hue are more easily classifiable, with if needed amplifications of distinctions by referring to brightness and saturation. So the shifting prominence of lightness and hue as the segmentation of perceptual colour space progresses does not invalidate the universalist thesis of a shared, pan-species neurophysiological substrate.

A shortcoming of IDM which will become relevant later is that it makes reasonably heavy use of cognitive representations. Jameson distinguishes:

"(1) individual perceptual representations (e.g., discrimination based); (2) individual cognitive representations (e.g., matching and tolerance based); and (3) a shared cultural representation (e.g., color lexicon based)." (Jameson 2005, page 31)

Jameson suggests that the mappings between these layered representations requires an additional cognitive layer, as a translation algorithm between the various domains. My suggestion, to be substantiated in chapter 7, would be that representation category (1), and at least part of (2) need not be representations in the cognitivist, computational sense. In accordance with Evan Thompson's Gibsonian inclination (see section 4.5), my suggestion will be that for those cases we can make do with ontologically less presumptuous entities. In chapter 7, I will offer an $E_{(i)}C$-*compatible* redefinition of the notion 'representation'.

5.3 - Synthesis

In the previous section, we saw that Kimberly Jameson provides us with a plausible characterisation of the progressive perceptuo-linguistic segmentation of colour space, in the form of her Interpoint Distance Model. Earlier, for instance section 4.5, it was suggested that we should adopt a theory that is a bit more subtle: categorization, both the linguistic type as applied to perceptual space, and concept formation, consists of an interaction of *biomechanical* (possibly innate) and *environmental* properties.

What might be innate (and, as such, underlying and structuring perceptual colour space), is the propensity of biological systems (agents) to develop, in interaction with the environment (with a particular illuminant distribution - see section 4.5), what I like to call a specific *neurophysiological yield*. This structure of saliencies *is* - at least in part - genetic in origin, but note that this does not imply the claim that the structure of percepual space for colour is exactly the same for all individuals in all cultures, as (the charicature of) an orthodox universalist might wish to claim. It *is* reasonably uncontroversial, however, that for the vast majority of humanity, the possible variation in idiosyncratic neurophysiological yields is limited to a narrow band. That is, it is likely that any 'normal' neurophysiological yield is sufficiently similar to another to enable language to yield a 'cognitive convergence zone': within a single language community, perceptual saliency structures do not vary to such an extent that talk about 'red' or 'green' cannot be intersubjective. In other words, any translation mechanism to smooth out occurrent differences in equi-dimensional cases (which Jameson suggests for trans-dimensional mapping scenarios - see previous section) is likely to be fairly low-key, since there is limited need for such translations. One reason for this is that colour language does not cut perceptual space nearly as finely as would be necessary to make most perceptual differences apparent in everyday interaction. Any occurrent cultural differences can be explained as differences in interpretation overlying a perception base that is largely convergent across members of various cultures, and these divergences (that are nonetheless *non-trivial*) constitute a more natural playing field for cognitive activity.

Cognition does have a role to play, mainly in the socio-cultural arena. For Jameson, colour categorization is both perceptual and cognitive, but one point of criticism about Jameson's IDM can be that it unjustly downplays the role of physiologically determined perceptual saliences of specific colour exemplars in favour of the hypothetical possibility of widespread divergences in chromatic receptor dimensionality. Still, a major advantage of allowing the influence of the cognitive domain into a quasi-universalist account of colour categorisaton, would be that the cultural dimension Saunders (1992) emphasizes might be incorporated as well. She noted for some cultures, colour words do not exist in abstracta, but rather were always linked to culturally significant objects, plants, animals, and so on. If the cognitive processing layer is added, as Jameson suggests, cases of (say) a tetrachromat asking herself how a variety of colours she gets to distinguish might fit into that single category that is used by the majority, could perhaps fairly easily be generalised into an account where there is room for a semantic cross-pollination of a specific shade of green, the plant that has that particular colour, and the word used to refer to the plant and/or the colour and/or the cultural use of the plant.

In chapter 6 and beyond, I will develop an account about the relation of perception and cognition (or rather, an account of cognition and concepts as an aspect of an agent-environment perception-action dynamic) that will

be compatible with the muted Gibsonianism in my treatment of $E_{(i)}C$, e.g. as evident in Thompson's enactivist/ecological colour theory.

I wish to claim that $E_{(i)}C$ necessitates acknowledging all modifying influences - be they body-based, environmental, cultural, and even cognitive. The inclusion of this latter aspect (cognition) need not take the form of a conscious processing of information that is in itself user-neutral (which would be anti-Gibsonian): in line with the theory of affordances, the animal can already be predisposed to process the incoming data in a specific way, according to its needs and the way it is embedded in the environment. So, while an animal might not possess linguistic categorization abilities, its specific being-in-its-environment, including those features of that environment it is instinctually inclined towards utilising or avoiding, embodies a primitive version of the suite of *constraints* upon colour vision not intrinsic to the physiological vision system itself that, in the case of humans, gives rise to highly complex cultural structures, language included.

Returning, for the moment, to the colour language case (as a less complex example for the kind of scheme necessary to explain concepts in $E_{(i)}C$-terms), my suggestion is (as stated before) to stake out a middle ground between universalism and relativism about the colour-centric interaction of agent and socio-cultural environment. This is not unlike the way I developed a hybrid account from the opposition between Thompson and Shepard concerning the colour-centric interaction of agent and *physical* environment.

For the socio-cultural case, I claim that, despite variations in external influences (environmental, socio-cultural, linguistic, and so on), an appropriate theory should be able to account for the fact there are also substantial similarities between individuals humans within and across cultures.

This dynamic of divergence and convergence might be explained by the following theses:

(1) the need to categorise itself is a practice enforced by the complexity of an agent's embeddedness in and his interaction with his physical environment;
(2) the specific character of categorization is:
- (2.1) in some ways *universal across cultures* by virtue of the fact that:
  - (2.1.1) all humans possess similar neurophysiological visual systems;
  - (2.1.2) members of those various cultures, despite their differences, might run into very similar problems (limitations and potentialities having to do with the way the human body reacts to specific circumstances, for instance), which categorization strategies are designed to help solve;

(2.2) in part *communal, but culturally specific* by virtue of shared language and customs amongst members of a culture;

(2.2.1) this can help shape a 'categorization-convergence zone' for neurophysiologically divergent subjects:

(2.2.1.1) in intra-cultural cases in the form of a translation-mechanism on the subpersonal level, to smooth out any occurrent differences in perceptually equi-dimensional cases (e.g. similar to what Jameson suggests for trans-dimensional mapping scenarios);

(2.2.1.2) in inter-language or inter-cultural cases in a more overt, at least initially conscious linguistic and/or cultural translation mechanism (e.g. <English> 'red' = <German> 'rot' = <Dutch> 'rood').

Based on these theses, I can now make good on my claim (from chapter 4) that a good theory of perceptual colour space segmentation should take cues from both the relativist and universalist traditions. Finding that middle ground between universalism and relativism entails claiming that *both* pan-species physiological properties *and* contextual influences (physical and socio-cultural) are of relevance. This means that the following components contribute to a *context-dependent* dynamic of constraints and enablings, yielding a specific colour-involving relation of an agent with his environment:

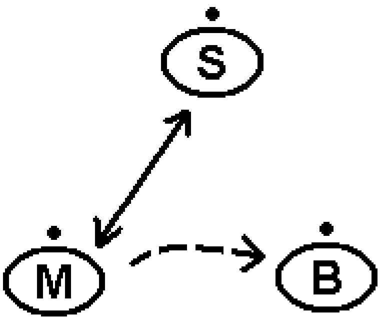
(1) *the neurophysiological yield*: the structure of hue / brightness / saturation saliencies determined by the properties of the physico-sensory substratum (which involves the interrelation of the physical processes producing the stimulus, and the neurophysiological mechanisms that process the stimulus);

(2) *environmental prominence* (e.g. there will be more pressure on an organism to devise a proper categorization scheme for various flavours of green if the difference between those shades is important to its continued survival);

(3) *socio-cultural prominence* (as an addition to / modulation of environmental prominence; e.g. the use of artifacts or agricultural products of a specific hue might be significant in a religious ritual, which might imbue the associated colour names and concepts with a meaning related to the ritual, and/or anchour the associated colour name at a position in the hue name acquisition hierarchy that diverges from the one in the standard Berlin and Kay evolutionary sequence). Socio-cultural prominence includes *linguistic encoding saliency*: properties of the language itself might influence category and/or concept formation; perhaps I can call these *Sapir-Whorf-effects*.

Compared to the general scheme of physical/ecological situatedness presented in section 4.5, the discussion above adds other influential factors. In particular, we have seen that convention (socio-cultural prominence)

influences the ways in which physiological properties (the neurophysiological yield) contribute to a specific kind of colour-related behaviour. In a schema already shown (by way of a preview) in section 4.6, this looks as follows:



[Figure 8: interrelatedness of agent and *social* world]

To repeat the description given then, this diagram depicts the following: the biomechanical properties of the agent as they change over time (dM/dt), in *interaction* with socio-culturally determined properties of the environment as they change over time (dS/dt), can, at some level of detail, be *described* as behaviour (i.e. the change of behavioural patterns over time, dB/dt). This behavioural description can be expressed in terms of a behavioural space, as a higher-dimensional version of the movement planning field utilised by Thelen et al. (2001). This past chapter can be understood as an attempt to provide a more detailed qualitative specification of how these agent-properties and *social* environment-properties hang together, at least in the case of colour perception, expanding upon the ideas about colour language in section 4.3.

In chapter 8, these two schemata will be combined into a schema depicting certain properties of concepts as defined in an $E_{(i)}C$-appropriate fashion. In the description above, we can already see the additional element that will be added: 'environmental prominence' will be accounted for in terms of the properties of the physical environment.

Based on ideas about the linguistic categorization of perceptual colour space described above, I will continue with an account of *concepts*, as well as cognition in general, but compatible with $E_{(i)}C$ - this will include ideas about *conceptual space* as a spectrum ranging from base physiology all the way 'up' to abstract cognition.

= =

[SUMMARY of chapter 5 AND PREVIEW]

Based on the ideas of Kimberly Jameson, this chapter provided a more detailed exploration of the progressive segmentation of colour space. Starting from the most basic dark vs. light distinction, the distribution of

neurophysiologically defined colour saliencies in interaction with sociocultural demands determines the sequence of the most informative colour space segmentations. That is, the way in which colours are named and conceptualised occurs under influence of *both* ecological/environmental *and* socio-cultural factors (i.e. the prominence of coloured objects in, respectively, ecological niche and sociocultural practice).

In line with this, the example of the Hanunóo of the Phillippines - their colour words have complex non-hue correlations - suggests that the three-dimensional phenomenal colour space of the received view of colour perception should be expanded with additional semantic connections (where a colour concept denotes not just a hue, but also, for instance, a particular object or ritual, with their associated meanings); this is the first step towards a proper conceptual space as a higher-dimensional version of perceptual space.

Jameson's Interpoint Distance Model prepares the way for an account of conceptual space segmentation, involving cross-dimensional mapping as a mechanism to relate complex segmentations to more basic structures (and the other way around - see also section 6.9 as this idea is linked to George Lakoff's ideas about metaphor-based concept development), and more in general the interplay of the agent's body-based, social-environmental and physical environmental properties.

In chapter 6, a more detailed account of concepts as behavioural dispositions to fit in with this 'conceptual space'-idea will be developed; in chapter 7 and beyond the interrelatedness of conceptual space (C-space) with M-, S- and P-space (respectively: bodily properties, sociocultural environmental properties and physical environmental properties) will be explored.

= =

[6 - Superposition Theory of Complex Concepts]

6.1 - The 'Colour' Concept

In the sections to follow, I will formulate a theory of concepts, traditionally the basic buildings blocks of thought processes. More specifically, I will start by suggesting a model to account for a specific *complex* concept, namely the concept of 'colour', that can be generalised to a model of concepts proper. This model, in turn, forms the initial building block of the 'Radicality Manifold'-model I take to fulfill the promise made in the introduction to this book: it will help position concepts within the embodied and embedded cognition paradigm. Recall, from my remark in section 4.1, that the real focus of this book is to develop an insight into the concept 'concept', not necessarily into scientific concepts, even though the current discussion will contribute ideas about (the components of) the scientific concept 'colour'.

The main problem with providing a straightforward ontology associated with the concept 'colour', is that the colour vision process cuts through several different domains of scientific investigation. The properties of coloured objects, as well as the electromagnetic radiation that is indispensible to the whole process (i.e. light), is best described using theories of (micro-) physics, which in this case will include quantum mechanics and optics (see e.g. Nassau, 2001). The sensory processing involved in colour perception can be described by neurophysiological theories, but also in terms of psychophysical models (see e.g. Hardin, 1988). The problem is that the claims of the theories associated with sensory processing are not necessarily expressible in terms of the theories that are most appropriate to describe the (physical) properties of coloured objects; the realm of colour perception comprises even more of such domains governed by mutually irreducible explanatory strategies.

With a concept that has so many different uses and applications, the kind of work you want the concept to do determines much of its content. Recall the claim, made above, that the phenomenon of 'colour' cuts across various disciplines and theories; this implies that the kinds of questions that one asks, i.e. the practical goals and purposes of (scientific) inquiry, determine what kind of explanatory framework (theory) needs to be utilised. Each relevant theory, then, has its own way of specifying how to characterize colour, and each of these specifications exerts its own force on what colour as an abstract notion would have to mean (if this abstraction is at all possible within the framework in question).

Here are some examples of disciplines that are relevant to explaining (certain aspects of) the concept 'colour', followed by examples of the kinds of phenomena or properties that have been suggested as candidates for ontological identification and/or the kind of role that 'colour' would occupy in each of these disciplines.

-**Physics**: colour is a microphysical structure, or a feature of light that is constituted by properties of electromagnetic radiation;
-**Linguistic Anthropology**: colour is a socioculturally relevant perceptual feature, that is imbued with relevance and meaning in a partly contingent, culturally determined fashion;
-**Phenomenology**: colour involves qualia;
-**Psychophysics**: colour is a structured percept, with propreties that depend, to an important extent, on the properties of the neurophysiological structures involved in processing visual stimuli;
-**Neurophysiology**: colour is input or information[NOTE 33] that needs to be processed;
-**Everyday parlance**: colour is an object property with certain informational properties[NOTE 34].

The problem is as follows: it is difficult to see how something can be defined as a physical structure, a meaningful perceptual feature, a phenomenal 'feel', an information-bearing signal and several other things all at once, without a significant loss in contextually relevant information. Despite this great variance in context-related applications, the everyday use of 'colour' as a phenomenally indexed object property appears unproblematic, and that is somewhat puzzling.

6.2 - Complex Concepts: Preliminaries

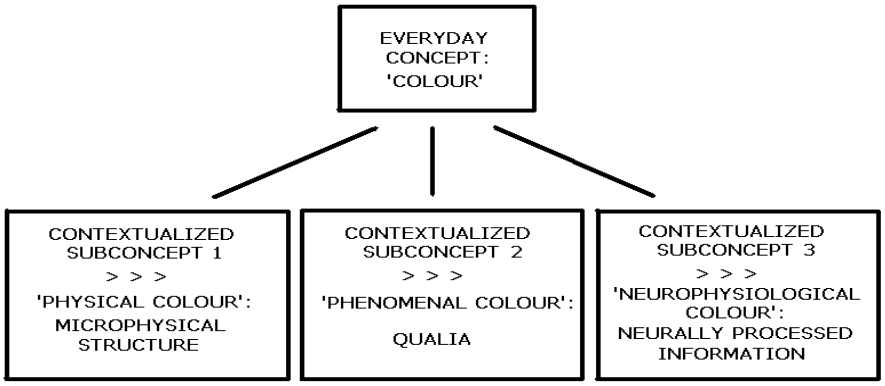
To account for the features of the concept 'colour' as described above, I submit that colour is a *complex concept*. This means that there is a tension between the apparent singular meaning of the concept as it is naively used in normal language, and the wide array of possible actual meanings that is revealed when the concept is called upon to account for some phenomenon falling within its application-domain[NOTE 35]. That is, when the general concept 'colour' is applied in a specific context, it starts to mean something subtly different, for instance what the scientific discipline involved in explaining phenomena in that particular context prescribes that 'colour' should mean. Such a contextualization of 'colour' involves hiding several other possible applications of 'colour' - several other *subconcepts* - from view.

It still pays to view 'colour' as a single concept, at a low level of detail anyway, because of the apparent ease of switching between subconcepts, each appropriate in its own domain of application. When we speak of colour in a phenomenal context, it usually does not appear to be too big a stretch to consider talk about colour in a physical context as talk belonging to *the same concept*, despite the incommensurability of the theories associated with each domain. So we have one concept, 'colour', which comprises a collection of different subconcepts, each with its own practical application (namely as specified by a particular theory or practice; see figure 9).

A complicating factor is that the ontology of the phenomenon is usually determined (or at least *investigated*) on these 'subconceptual' levels, hence

the structures and contents of the theories associated with the subconcepts are very influential in the way concepts are defined, understood and used. This results in the odd situation that there are different ontologies associated with what, in everyday parlance, appears to be a perfectly straightforwardly definable property ("'Colour' is that property *right there* on the object!").

In brief, the idea of a complex concept implies that there is not merely a peaceful division of labour between the various subconcepts, but that there are incommensurable stories intended to refer to or somehow subsume under the same notion - see figure 9. In the case of colour, this would be the notion of a phenomenally indexed object property.



[Figure 9: various subconcepts of the naive concept 'colour']

I wish to contend that colour is not the only concept that displays this structure - in fact, there might actually be many complex concepts: good examples are 'information' (which could refer to semantics, syntax, signals or stimuli, and any one of a wide array of socio-cultural variations on these themes) and 'time' (experienced time has rather different properties than time in physics, and even within physics there are different possible ontologies, e.g. absolute/substantivalist vs. relational - see e.g. Rynasiewicz (1996)). Whether there are many or relatively few complex concepts does not actually matter all that much: what matters is that at least *some* important, salient, widely used concepts have this property, because that is enough for it to be a phenomenon that theories about concepts will have to be able to account for. The theory about concepts to be described in the next section, Superposition Theory of Complex Concepts, will actually be a theory about concepts in general, which will include complex concepts as a special case.

The most important criterion for a concept to be complex is that the various definitions of the overarching concept should not be merely metaphorical variations on a single theme. For instance, a case can be made for the idea that the pain I might feel due to an unrequited love can be thought of as a metaphorical extension of the kind of pain I can feel in my arm. In contrast,

a concept is complex if the properties associated with its various subconcepts perform distinct, mutually irreducible roles. This is the case if these subconcepts are to be explicated in terms of the vernacular provided by different, irreconcilable theories. This is why 'water', for instance, is not a complex concept, even though there are different ways of using the word that do not automatically evoke the same content: chemists, ship engineers or athletes on a hot summer day can seem to mean rather different things if they speak of 'water', but it appears likely that all of the properties of the substance in question (e.g. having a particular boiling point under certain atmospheric conditions, the power to facilitate buoyancy or the power to quench thirst) that are relevant to each of these categories of water-users can quite comfortably be explained by referring to the chemical properties of the H_2O-molecule, and the physical properties of aggregates of these[NOTE 36].

But when a concept *is* complex, its complexity is usually due to the fragmentation of science, and/or the (to some extent parallel) fragmentation of everyday parlance. As such, the existence of complex concepts exposes the holes in the fabric of our understanding and description (scientific or otherwise) of the world - holes in want of a patch, or an entirely new cloth. Still, this predilection towards multi-denotational ascription does appear to be part and parcel of some and perhaps even much of our concept-use, meaning that any theory of concepts should be able to explain how we do so, and why this is the case.

Despite the fact that not all concepts might be complex in the sense explained above, there is another important consequence to be gleaned here, especially in light of the discussion of previous chapters. Complex concepts are context-dependent to a very high degree, but the $E_{(i)}C$-focused discussion of behaviour and perception so far suggests that concepts and cognition in general are context-dependent, in the sense that their structure and content is informed by bodily, social-environmental and physical-environmental properties. The sections to come are dedicated to describing what kind of a conceptual structure is implemented because of those influences.

6.3 - An $E_{(i)}C$-approach to Concepts

As we have seen (in chapter 2), most theories define concepts as building blocks of thoughts, with a particular internal structure and content, attribute to them an important role in acts of categorization, and might invoke them to explain the systematicity and productivity of thought. In other words, concepts are mostly or almost entirely entities that need to be defined in terms of cognitive processes: concepts are mental entities.

In $E_{(i)}C$-theories, the lines between cognition and action are blurred. Bodily action is not necessarily *controlled* by cognitive processing, but might be structured, to an important extent, by bodily dynamics; cognition is influenced by (subconscious) sensorimotor processes, which depend crucially on properties of the body and the perceptual system; and cognition

can utilise aspects of the environment to support its own developmental dynamics. This means that in $E_{(i)}C$, the kinds of processes that would require (or can be described in terms of) cognition, need not be exclusively mental - that is, if 'mental' is defined as in the head or the mind, and not (at least in part) constituted by the dynamics of the body and/or the environment.

Now, the standard claim about concepts, namely that they are 'constituents of thoughts', is somewhat opaque. It should be possible to make some headway on devising an $E_{(i)}C$-approach to concepts by making the link between concepts and the network of bodily, cognitive and behavioural processes, and the environmental properties and processes that these are situated in, a bit more explicit. One way of doing this is saying that having a concept is to be defined in terms of *achievement*, of being able to *do* something, of expressing a particular level of *expertise in a specific context*. This means a concept is no longer a mental entity in the classic sense (i.e. non-physical), but it can still be part of a *cognitive process* if the definition of the term 'cognitive' changes in accordance with the paradigm shift the $E_{(i)}C$-approach as a whole represents: a concept is still a constituent of thoughts, but what 'thoughts' means has changed.

In this sense, my suggestion is that a theory of concepts should take a broadly *Wittgensteinian* tack. That is, in line with the inclinations of the $E_{(i)}C$-approach in general, and more in particular the $E_{(A)}C$-approach, cognition, in its essence, should be understood to involve body-based and environmentally situated activity, rather than stacks of representational layers all the way down to the level of basic mental processing. However, it is important to note that I do not intend to get rid of representations altogether: the very essence of the accounts of concepts and cognition to be developed in the pages to follow rests on the notion that the usual $E_{(i)}C$-approaches fail to account for higher cognition, and in a nontrivial subset of the cases that $E_{(i)}C$ cannot account for, things like Andy Clark's (1997) 'representation-hungry problems' are in play. However, this use of representation turns it into what can be called an additional layer of abstract processing that, in some agents, is added to a more fundamental, $E_{(i)}C$-style process of concept-involving behaviour. Furthermore, the notion 'representation' itself will be subject to some scrutiny, in chapter 7 in particular.

Given all this, it is possible to formulate the following general $E_{(i)}C$-definition of what it means to have a concept:

having a concept A of some object/process/state of affairs O means being able to act in an appropriate manner, given the possibilities P for and constraints on action CA that O represents, and given additional contextual constraints CC.

Most of the components of the definition above - 'object/process/state of affairs', 'being able to act in an appropriate manner', 'possibilities for and

contraints on action', 'contextual constraints' - allow for a fairly natural specification in embodied and embedded terms; at the very least, they depict processes and entities that derive their properties at least in part from the position they occupy within a broader context, and the model to be developed in the remainder of this book will investigate what that position and its context are like.

I would imagine the 'appropriateness'-criterion in the provisional definition above raises the greatest number of eyebrows. As a brief preview of discussions to come, it can be said that this *normative* aspect is defined (to an important extent) by the web of *social affordances* constituted by the actions of conspecifics, in addition to the *physical* norms laid down in the 'regular' affordances (possibilities for agent-to-object interaction; see sections 4.5 and 8.2).

The components mentioned in the definition above carve out the content of the concept 'concept' collectively, and this content so far remains as an unknown waiting to be filled in. A provisional description of 'concept' that I believe will fit is the following:

a concept is a structured behavioural disposition of an embodied and embedded agent.*

It should be noted that, in this corollary, 'behaviour*' should be understood to include not only bodily action, but also locution and cognition. The idea here is that concept possession can be expressed in many ways, of which linguistic description provides important, but merely partial coverage. This all means that talk of 'having a concept' as if it were an independently describable entity is profligate; 'activating' a concept-as-disposition consists in 'activating' a capacity, which translates directly to 'acting'.

However, if concepts are defined or identified in terms of action, or dispositions for action, and the criterium 'efficiency' is included in the definition, that means that it is possible for an agent to be more or less adept at implementing conceptual knowledge and abilities, and this is, at least in part, a function of the conceptual content that is available[NOTE 37]. To put this another way, it should be possible, at least in principle, to specify a spectrum or hierarchy of concepts. The 'efficiency'-criterium of concept-informed action can be understood in different ways, and these different criteria yield different hierarchies, but one way is to parse efficiency in terms of notions that skew the balance towards the kinds of (cognitive) abilities humans are comparatively good at, such as *creativity*, *versatility* and *adaptability*. If this is done, one possible way to construct this hierarchy is to have it range from the most basic action- and perception-based concepts (or concept-like abilities and dispositions) at the bottom, up to high cognition, including abstract thought and creative imagination.

This way, conceptual ability is measured in terms of cognitive capacity, and of course this yields the somewhat familiar line along which agents of all

kinds can be grouped, with primitive animals near the bottom, and dolphins, chimpanzees and humans near the top. Because this is just one way of organizing this hierarchy - other ways of defining 'efficiency' are likely to result in rather different rankings, with humans nowhere near the top of the heap - I would argue against attaching too much importance to it. However, I *would* like to argue that this particular organization, ranging from basic sensorimotor abilities all the way up to high cognition, does offer an appropriate template for the organization and functioning of the human conceptual system. SToCC, and the model of $E_{(i)}C$ to be developed on the foundations formed by SToCC, will make some use of this idea; furthermore, the claims involving the structuredness of the 'colour'-concept (and other complex concepts) presented in sections 6.1 and 6.2, will also be integrated into this account.

But first: to introduce the idea of this spectrum of dispositions, of which our familiar ideas of what concepts are (and who has them) form a limit case (namely, the upper limit), is to present a theory that, in some sense, yields a deflationary view of concepts[NOTE 38]. In lower regions of this spectrum, there might be abilities and dispositions that are concept-like, but not actually concepts themselves if we understand 'concept' to require consciousness and higher cognitive abilities (including the capacity for language). However, SToCC will imply that the familiar kinds concepts (i.e. as they are used by humans, involving mostly linguistic categorizational abilities) do not form a clearly demarcated island in the ocean of non-conceptual processing, but that there is a smooth continuity between these high-end concepts and the lower-end somatic abilities and dispositions[NOTE 39]. SToCC makes a claim that is stronger still: many of the 'higher' concepts depend on lower concepts, and are often partly constituted by them. How this can be the case will become clear in the sections to follow[NOTE 40].

Another way in which the current suggestion can be seen as a deflationary account of concepts, is more radical. That is, it would even be possible to go so far as to say that in acquiring concepts or judging whether someone else has the same concept we do, the relevant process *does not involve* copying and comparing concepts as such, but copying and comparing *behavioural profiles* that we can use as *indicators* of concept possession. This would mean that the term 'concept' is merely an abstract description of certain structural elements of a disposition towards a particular behavioural profile. In other words, a 'concept' (in general) is nothing in and of itself, but an abstract description of a concrete act, or disposition towards it.

It might seem as if this view defines concepts merely in external, perhaps almost behaviouristic terms, thus ignoring the essential *internal* aspects of concepts and concept possession: at the very least, having a concept also means having specific knowledge, right? I can assure the reader that these aspects will not be forgotten; see the remainder of this chapter for more about the way SToCC views the content and interrelatedness of concepts, sections 6.6, 6.7 and 6.8 in particular.

It is at this point that I wish to add a critical note about the concept 'concept'. There appears to be a tension between two opposing forces, each pulling the definition of 'concept' in its own direction. On one side, we have language, which allows us to apply neat, clear-cut labels to all kinds of objects, processes and abstracta; a strong enough emphasis on this pole will make us think of concepts as transparent, uniquely referential categorizations of the world. On the other side, we have the everyday hustle and bustle in which concepts function, which is a complicated dynamic of processes and forces; emphasis on this pole will make us see that concepts are overwhelmingly multiply realizable, not only in terms of the idiosyncracy involved in the specification of a concept's internal structure, but also in terms of the multitude of available variations (*intra*personal as well as *inter*personal) in behavioural patterns that can bring about a particular goal, hence can be linked to a single concept (or close-knit group of related concepts). Let's call the situation that results from the presence of these two opposing forces *the inherent instability of the 'concept'-concept*.

From this idea we can extract an important implication: a theory of concepts will have to be able to address the issue of how we can speak sensibly of two people possessing the same concept in the face of this multiple realizability. Recall that this was one of Prinz' desiderata on a theory of concepts (see section 2.5). The unifying force of language is one possible explanation that is available, but a more detailed description of the mechanism at work here, or possibly suggestions for an additional explanatory mechanism, will be most welcome. SToCC will offer a few suggestions of that kind, including mechanisms involving 'conceptual enslavement' (see section 6.7) and 'granularity' (see section 6.8).

6.4 - Conceptual Space

Central to StoCC is the notion of 'conceptual space'[NOTE 41]. This is a metaphorical notion, and it is introduced here as an *extrapolation* of the perceptual space used in, for instance, the received view about colour perception as described in section 4.1. Conceptual Space involves the idea that concepts as behavioural dispositions are structured in a particular way, namely according the individual concepts' inferred accounts. That is, an important part of having a concept involves being able to exhibit some kind of behaviour or provide some kind of explanation that is accepted by others as instantiating the appropriate kind of justification for that particular use of that concept, and this justification occurs along inferential lines. Conceptual space is a metaphorical way of understanding the interrelatedness of concepts that results from these justificatory connections. See section 6.6 for a more elaborate explanation - the point I wish to make now is that at the foundation, such justifications often depend on the biological particulars of our embodied, embedded interaction with our environment.

From the discussion in chapters 4 and 5 about the specifics of colour-centric embodied and embedded agent-environment interaction, we can take away a notion of what the internal dynamics of conceptual space could

be like. That is, there are important connections between the basic perceptual topology (e.g. the fact my eyes and brain are configured in such a way that I consider a particular shade of red 'the best' red) and higher-order semantic contents (e.g. my predilection to classify objects based on their colour, and the behavioural consequences this has, for instance in front of a traffic light).

This idea aligns quite nicely with that of a spectrum of concept-dispositions, as introduced above (in section 6.1), as well as the idea of the lower end of that spectrum blending in smoothly with the perceptual / somatic / nonconceptual realm: perceptual space acts as the basis of this proto-conceptual space, and an increase in complexity is, in principle, quite easily expressed by an increase in dimensionality, with each of the additional dimensions (or coherent set thereof) expressing some property with which to specify some concept.

For now, the basic claim is that the idea of 'conceptual space' entails that it is possible to define a space that comprises the conceptual complexity spectrum, and is ultimately rooted in basic sensorimotor activity. This should be enough, at least for the moment; however, in sections 6.9 and 6.10, I will say more about the structure of conceptual space, and the way in which its 'higher' reaches depend on/relate to the basic sensorimotor level, and to the various aspects of the agent-environment interaction dynamic. Before these hypotheses can be given, more about the properties of conceptual space, and the concept-user's embeddedness in the environment (described in terms of a broader structure called the Radicality Manifold), needs to be said.

Towards that end, I will devote the subsections to follow to highlighting some properties of conceptual space. In particular, I will explain how complex concepts fit into this framework, namely via an internal structure of concepts, subconcepts and *conceptual superposition* (section 6.5), what role *inference* plays in determining the structure of conceptual space (section 6.6), that some aspects of concepts, *enslavers*, are more important than others (section 6.7), how *granularity* is an operator relating conceptual space to the agent's context of action (section 6.8), how the evolution of concepts might come about, for instance via *conceptual splitting* (sections 6.9 and 6.10).

6.5 - Conceptual Superposition

In section 6.1, I provided a (non-exhaustive) list with different versions of the concept 'colour' (as used in physics, linguistic anthropology, phenomenology, psychophysics, neurophysiology and everyday parlance), and argued, on the basis of that list, that the concept in question should be thought of as *complex*. Still, in everyday parlance, we do not appear to encounter many problems in using the single notion 'colour' to stand in for much more detailed accounts, even though some of those detailed accounts appear to denote entities or properties that cannot be captured in

terms of the kinds of explanations and descriptons that are provided by other relevant accounts.

To capture this property of a complex concept - a singular concept at the everyday 'level' that breaks apart into a variety of mutually incompatible accounts if placed under scrutiny (see section 6.1) - I want to introduce the theoretical description *conceptual superposition*.

What the idea of conceptual superposition says, in essence, is that at least some concepts as they are used in everyday parlance cannot be given a straightforward explication or definition - that they, in fact, yield *mutually exclusive* inferences, depending on the context in which they are supposed to apply.

The notion *conceptual superposition* is meant to invoke association with *quantum superposition*, which denotes the possibility that an object possesses two (or more) values for a particular unmeasured variable at the same time (e.g. the energy of an elementary particle). The moment this particular value is measured, the particle's wave function collapses into a determinate value. The depiction of this property occurs in terms of the addition of *state vectors*, and this method of description, plus the occurrence of wave function collapse, offers a good (at least metaphorical) fit with the property of a complex concept.

That is, first: SToCC promotes the idea of concepts as collectively specifying a 'conceptual space', which affords a description of concept properties and contents in terms of a *vector space* (at least in principle). And second: SToCC claims that in its naive, everyday form, a complex concept is a singular notion, a *conceptual superposition* that consists of the addition of several *higher-grained* but mutually exclusive subconcepts. When this concept is placed under scrutiny, it is applied in a specific field, or someone or something demands of us in some way that we articulate what we mean by this naive concept - i.e. the concept's properties are 'measured' -, the superposed concept breaks apart and reduces to a particular subconcept, that is tailored to the context in which it is supposed to do its work, but might no longer apply to different contexts of use, where other subconcepts of the superposed concept are relevant.

For instance, if a particular industrial application requires determining whether some object is light or dark under certain, precisely specified lighting conditions, the subconcept of 'colour' utilised here will probably be defined in terms of the wavelengths of radiation within the visible spectrum emitted or reflected by the object, because that is the kind of feature a machine to perform the above-described task is likely to measure. Picking this definition of colour means putting aside (at least for the moment) other aspects of the 'colour'-concept, such as the socio-linguistic and phenomenological aspects.

One way of making the introduction of superposition as a step in the development of a theory of concepts somewhat more tractable is by exploring this property in terms of *non-functionalizability*.

David Lewis (1972) offers a formal account of the definition of theoretical terms, which can include terms denoting mental states. As such, his article was one of the formative contributions towards the theory in philosophy of mind called *analytic functionalism*. I do not plan to discuss or defend Lewis' analytic functionalism in a detailed fashion, but functionalizability (or the impossibility thereof) is a useful notion in the current context.

This is a cursory description of how Lewis develops his idea: suppose we are uncertain about the ontology that underlies a particular state (or process or entity), but we do have a body of peripheral knowledge, that collectively specifies a particular role for that state to perform. Lewis suggests we recast the story about this state as a conjunctive sentence, composed of T-terms (theoretical terms, the terms to be defined and explained) and O-terms (other terms; old, familiar knowledge, e.g. folk-psychology, in the case of a story about mental terms).

The T-terms denote roles, the properties of those roles are specified by the O-terms, and the realizers occupy those roles. That is, real-world entities, if in fact the story is true about these (and only these) entities, are said to *(uniquely) realize* the theory. If a story is incorrect on a detail, this would imply there are no realizers to be found to occupy the role specified by the (incorrect) theory. However, in that case the (incorrect) story is *nearly realized*. - the T-terms name the components of a *near-realization* of the corrected form of the story (i.e. as it should have been told).

As said above, if the states to be explained are mental states, Lewis suggests we take folk-psychological statements as O-terms, and form a conjunction of them, to specify (more or less implicitly) the kinds of roles mental terms are supposed to play, hence what they should be defined to be, in terms of our everyday way of speaking about them. More explicit definitions of T-terms (mental states, in this case) are formed by formulating the causal relations in which mental states stand. If we find out what kind of phenomenon performs the appropriate role, given that set of causal relations, we will know what the mental state in question *is*.

If we wish to explain what a particular concept means, we can do at least part of that work by developing Lewis' scenario in the opposite direction. That is, we can specify the kind of role(s) a concept plays by localising it in a network of implications and inferences - we can reconstruct the concept's relevant subregion of conceptual space (see below, section 6.6).

The most distinguishing feature of the naive colour concept (and, as I would want to claim, every other complex concept) is that it resists exactly this kind of a manoeuvre: a complete list of truths associated with this concept will contain contradictions and incompatibilities. The claim is that for a

complex concept, it is impossible to find a realizer, or even near-realizer, for the suite of O-terms that exists about it. In other words: what is 'it', then? Recall the list of colour-subconcepts given in the previous section: they can not all be true of the same phenomenon or entity at the same time, yet when we use the everyday concept of colour, it somehow affords smooth inference to any of the subconcepts mentioned - in some peculiar way, we are comfortable with using a single (superposed) concept as a placeholder for that wide array of mutually exclusive subconcepts.

It might be possible to provide at least part of the explanation why this is the case. The idea that a superposed concept should still be treated as a single concept, despite the incompatibilities between its subconcepts, depends on two criteria:

(1) *Etymological*: the presence of a unitary notion or concept as the *historical origin* of the suite of subconcepts, which was refined and fractured in a process of conceptual splitting (see section 6.10 for more on this);

(2) *Practical*: the ease and familiarity with which we can switch between using the various subconcepts, and still have the strong intuition that we are talking about the same thing.

These switches between superposed concept and subconcept, or between subconcepts, is governed by the concept's internal structure. This structure is determined by a given (sub-)concept's *inferred account*.

6.6 - Inferred Accounts and Narratives

Recall once more that in section 6.1, the complexity of the concept 'colour' was illustrated by a list of descriptions of the ways in which various scientific disciplines explain and define colour. This implies that a particular concept, or the subconcepts belonging to a particular superposed concept, can be specified in terms of its associated theory. So in order to define what we might mean with the subconcept 'physical colour', we could try to formulate a description in terms of the laws, definitions, relations and regularities made available by physics, i.e. surface spectral reflectance, electron-photon interactions, energy bands, and the other concepts and theories that describe the physics of colour (Nassau, 2001). Of course, each of the concepts used in these theories has a structure of its own, and can be explained in terms of other theories (possibly more basic ones, in a mereological sense), which are likely to imply still other theories, and so on.

Picking this method to define concepts would imply two things. First, it is possible to specify a *layered structure* of conceptual space; the 'superposed concept to subconcept'-transition is but the first step in a long chain of ever more fine-grained, basic definitions. In addition to these vertical connections, there are likely to be many other kinds of connections as well: some concepts might be linked to (i.e. play a role in constituting or explaining) several different other concepts simultaneously. This results in a

layered, web-like structure for conceptual space, with specific regions denoting a particular concept used[NOTE 42]. Each region (concept) has a structure, and at a higher level of detail, still other structures might reveal themselves: subconcepts of the concept, and 'sub-subconcepts' belonging to the subconcept, and so on. Hence, we can say that conceptual space consists of *multiply embedded manifolds*.

The superposition-relation (characteristically implying the mutual exclusion of subconcepts) is a special relation, and will not occur at every intra-level transition, but the structure in which a higher-level concept contains several other, lower-level concepts is ubiquitous, meaning that at least some regions of conceptual space are fractal-like in structure[NOTE 43]. Noteworthy is that superposition need not be limited to just the top level (i.e. the region of conceptual space containing little definitional detail), but might occur wherever a concept straddles a nexus of different kinds of explanatory accounts.

The second implication of the chosen definitional method is that a concept, if used in a particular context, e.g. to play a role in the explanation of a particular phenomenon, *does* point towards all the other theoretical baggage that goes along with it, but this extra content does not need to be present or implied in the use of that concept in that context. In other words, in some cases it is perfectly fine to use a concept as a primitive notion. In such a case, I call that concept a *contextual primitive*. For instance, it is possible to specify a specific surface spectral reflectance profile, and have this be a perfectly acceptable explanation of why an object looks to have a certain colour (given a particular line of questioning or investigation), even though the very concept 'surface spectral reflectance' implies many more theoretical notions and relations.

However, the notion 'theory' with which to specify the content of a concept (i.e. an embedded manifold in conceptual space) is too restrictive: not every concept or subconcept has a neatly worked-out theory to define its meaning. In fact, the vast majority of concepts does not. SToCC bears some resemblance to Theory theory of concepts (see section 2.4), which does state that concepts are structured in terms of theories, but there are several important differences; the first to be discussed is they way in which concepts are to be defined[NOTE 44].

SToCC suggests a characterization of the relations between concepts and subconcepts in terms of *inferred accounts*. Such an 'inferred account', which fills in what a concept or subconcept means, is not necessarily a full-blown scientific theory - in fact, more often than not it will consist of an after-the-fact, possibly almost apologetic account. This is the backbone of the structure of conceptual space in the SToCC model: the relations between concepts are to be explained in terms of *accountability*, in terms of *justification* That is, someone can be said to possess a particular concept if he is able to explain what he means in a coherent manner, or act appropriately in some other way that demonstrates he grasps the concept

at some level of sophistication (e.g. perform a behaviourally expressed categorization task with some level of success).

The criteria to measure concept-possessing-behaviour against can then be defined in terms of *achievement*: is the concept-based action successful? Does it help the agent achieve whatever it is that he wanted to achieve, or does it at the very least enable a good enough attempt at reaching the desired goal-state? Are the discussion partners, to whom someone's use of a particular concept's content is explained, satisfied with the answer they have been given? This satisfaction might not take the form of agreement, but rather *respect*: the array of differences in opinion is probably finer-grained than the array of differences in concept, so in some cases we might wish to concede that we use the same concept, even though there is some difference in *the way that we use* said concept[NOTE 45].

It is possible to differentiate the 'achievement'-criterion for concept possession in terms of behavioural, cognitive and phenomenal aspects, each aspect yielding its own subcriteria, as follows:

(1) *Behavioural* aspect: does the agent exhibit the appropriate kinds of action? For instance, how effectively does the agent capitalise upon opportunities for action in the environment?
(2) *Cognitive* aspect: is the agent capable of creativity and devising contextually appropriate strategies? An additional (but not essential) criterium could be the agent's ability to provide a linguistic report on the *reasons* for his actions.
(3) *Phenomenal* aspect: are there good reasons to suppose the agent possesses some kind of structured phenomenal content? Clues to help make this assessment can be derived from knowledge about the agent's neurophysiological structure and sensory faculties, and inferences based on criteria (1) and (2).

The above, if coherent, means that *in general* concepts are not informed by theories, but by the broader phenomenon of *narratives*, i.e. the behavioural and cognitive 'jurisprudence' that we build up by living and acting, all the while using concepts, seeing other agents doing the same, and remembering effective behavioural profiles for application at a later date, in situations similar to the ones witnessed. This means that the normative aspect of this 'jurisprudence' is, to a large extent, defined in terms of the socio-cultural situatedness of agents. Many of these narratives are likely to be partly or wholly implicit, at least until an agent is asked to explain or justify his use of a particular concept, either in words, or in terms of action. And on this account there is no obstacle to the idea that some concepts *are* explicitly informed by theories - technical, scientific concepts for instance; these can be understood in terms of extraordinarily structured, shared, probably institutionalized, and explicitly defined and formulated narratives.

However, most of these narratives more than likely do not satisfy the rigourous criteria of logic and rationality that proper theories should conform

to: inquiries into the meaning of most concepts probably bottom out at some idiosyncratic dogmatic level just because that is how the agent learnt or experienced it. That is, the aforementioned idea of jurisprudence most often concerns subjective experience rather than objective data. It concerns the context from which the meaning of a particular concept derives, for a particular agent, i.e. the kinds of experiences said agent has had, in which this concept was forged. In the vast majority of cases, I wish to claim, such a context has a narrative structure - consisting of events following other events involving the agent himself interacting with other agents and/or the environment - and that context determines the content and character of the experiences and bits of knowledge that inform the concept in some nontrivial way. An African-American child being told about the ideas of Dr. Martin Luther King, Jr., and deriving some grasp of the concept 'justice' from those ideas, and the child of a Ku Klux Klan member hearing about those same ideas and constructing some notion of 'justice' from those stories, will end up having very different concepts of 'justice'. This means that many variables are of influence here: what the child already knows, who tells the story and why, and in what kind of socio-cultural context the lessons learned from those stories are to be implemented all determine how that particular concept takes shape in that child's actions and thoughts. All these variables indicate properties of the child's own life story: the ways in which the various players and settings of his own story relate to eachother co-determine how specific ideas and experiences are to be interpreted.

Implementing the concept that the child has acquired (e.g. behaving in a just manner, according to what he has learned), and experiencing the implications of that implementation, form a further contribution to the continued evolution of that concept: "last time I tried to act according to what I thought was just by telling the police officer that my father had indeed been speeding when we were pulled over. My father got a ticket and later got really angry with me, so apparently it is not always best to be completely honest...". Several such implementation experiences then come to constitute the jurisprudence associated with that concept, i.e. the memories and ideas - concrete cases of behaving in accordance with that concept - based upon which new implementations of said concept can occur.

This means that a concept can be said to have a narrative character in two ways: (1) the concept acquires the content it has by virtue of those constitutive experiences being embedded in a specific meaningful narrative structure, and (2) having been informed by those experiences in that particular structure, the concept contributes to the continued unfolding of the agent's own narrative.

This does mean that each individual agent probably has at least some utterly unique concepts, because the way in which this agent learnt the meaning of that concept, and the way in which he experienced those learning processes, are unlike the concept-constituting experiences someone else has had. However, even though at a sufficiently high level of detail different agents' concepts are distinct, it is entirely possible for those

agents to believe that they share a concept, simply because those detailed depths of conceptual meaning and etymology are rarely - if ever, for most concepts - explored (see section 6.8 for more on this idea). An important factor in this synchronization occurs because all speakers of a particular language use the same limited vocabulary to label concepts: subtly different idiosyncratic associations evoked by a particular word do not, in the vast majority of cases, prohibit the effective shared use of such a word.

The arguments for Dan Hutto (e.g. Gallagher and Hutto 2008) to introduce his *Narrative Practice Hypothesis* (as an alternative to simulation theory and theory theory in the 'theory of mind'-debate) run largely parallel to arguments for me to defend the narratives-account regarding the structure of conceptual space. Hutto's claim is that a structure of *practical* narratives embedded in everyday life, rather than knowledge of *theoretical* folk psychological laws, is the source of our ability to understand others as mental beings with mental states. These narratives are the stories, containing knowledge about reasons for acting, that were delivered to us in our upbringing, and continue to unfold for us in everyday interaction with others.

One of Hutto's main arguments against the claim that folk psychology underlies our theorising about the minds and reasons for acting of others is that folk psychology is not a proper theory, and that even if people construct predictions or explanations based on what could be called folk-psychological knowledge, this is not a deductive procedure involving general laws. Any and all regularities that might, in some cases, allow inference towards rules are extracted from narrative contexts. As noted before, SToCC differs from the theory theory (of concepts) for a similar reason: concepts, and the inferences that they support, very rarely take the form of proper scientific theories and *their* allowed inferences.

An additional similarity of the Narrative Practice Hypothesis and SToCC concerns the context in which these Folk Psychological Narratives come into play most prominently. That is, they are most explicit when something is amiss: we do not continually engage in active, explicit reconstruction of the lives of others in order to understand their moves and motives. Rather, when someone does something that baffles us, that is in direct conflict with what we feel he should have done, the need to provide a narrative justification of his actions is most pressing. Similarly, in SToCC, being able to appeal to conceptual jurisprudence (with its narrative structure) is a criterion for possessing a concept, but the situations in which you are called upon to make parts of this jurisprudence explicit will usually be characterised by a mismatch of some sort: I use a particular concept in a way that fails to align with the expectations of my discussion partner, and he demands an explanation.

There is a third parallel between SToCC, and the views of Hutto. However, this parallel has yet to be made explicit. In their (2008), Gallagher and Hutto augment the NPH with Gallagher's hypotheses about primary interaction.

Gallagher's claim is that a great deal of the way in which we interact with someone is informed by subconsciously perceived cues in body language, facial gestures and so on. This pre-conceptual co-attunement of agents can be seen as a precondition, and a continual source of input, for the interactive, narrative-driven practice of understanding others as mental beings. These two domains, namely 'intersubjective perceptual processes' and 'narrative competence' (augmented by a third, namely 'pragmatically contextualized comprehension') collectively cover the range of agent-to-agent interaction-processes, according to Gallagher and Hutto.

In SToCC, conceptual space is utilised as a tool to describe an agent's concept-guided action, and this space could conceivably play 'host' to the narrative and pragmatic components of Gallagher and Hutto's theory: the inferential connections that exist between concepts are specified in terms of pragmatically oriented, narratively informed accounts, that are (supposed to be) provided when a concept-user is pressed to explain his views and actions.

Given the above-mentioned components of the theory of Gallagher and Hutto, two things are still somewhat absent from the SToCC-model: a stronger, more explicit account of how concepts are embedded in a broader social context, and a description of the role and place of these intersubjective perceptual processes. Both these demands will be met in the development of the 'Radicality Manifold'-model, in chapters 7 and on.

The above helps us achieve four things:

(1) the concept 'concept' is dislodged from its strict connection to theories of the scientific kind[NOTE 46]. Obviously, not everything we use to explain concepts conforms to the (scientific) criteria of what a theory is supposed to be. However, the 'inferred account/narratives'-proposal retains the kind of specification of conceptual meaning and structure that is needed: the proposal suggests that the inference in question yields a structured account intended to express the interrelatedness of the elements and properties of a specific object, state-of-affairs or process with the purpose of providing a record, which can be available to personal or interpersonal access in contextual (situated) explanation-demanding inquiries. This structuredness depends, to a significant extent, on the structure of phenomenal experience: for instance, a particular decision by the court can *feel* wrong, because it does not align with whatever intuitive apprehension of what is and is not just was instilled in me via my experiences. This brings us very close to the realm of bodily and/or phenomenal feels, i.e. the structures prescribed by the phenomenal basis of conceptual space. This suggests that much of this narrative structure is likely to be remembered according to feelings, sensations and impressions that were once experienced by an agent, rather than fully formed theoretical accounts;

(2) the notion 'narrative' (as opposed to 'theory'), is allied more smoothly with the perspectival epistemology inherent in the embodied/embedded cognition-paradigm: a concept can be (and often is) an *idiosyncratic* entity;

(3) if mental phenomena require *diachronic*, rather than *synchronic* definitions, the notion 'narrative' offers a much better fit[NOTE 47].

(4) The reason why concepts are not necessarily linguistically mediated or even symbolic representations, or any other type of mental entity (even though they can be, in some cases) can be summarised by the slogan 'the world is its own best representation'. What gives a concept its meaning is the way in which it is *used*, and how it is used is determined by, and to a large extent *instantiated in*, the meaningful action of an agent in a structured environment. For humans, this environment includes other agents, and the socio-cultural and linguistic structures they generate. These structures provide the constraints within which the agent is supposed to be a concept-user - that is, these structures provide tools and tutorials he can use to hone his concept-using skills, hence they provide the criteria against which his efficiency in concept-use can be measured.

To reiterate, conceptual space can be seen as an *embedded manifold*, a space containing structures within structures, in which the relations *between* structures and *internal to* structures are specified in terms of inferred accounts; these accounts are derived from the narrative formed by the agent's own experiences. There are two special characteristics of conceptual space that need to be spelled out a bit more: *conceptual enslavement*, specifying the internal structure of a concept, to be discussed in the next section (6.7), and *granularity*, an operator that defines which aspect of a concept (e.g. which 'layer') is active, to be described in the section after that (6.8). These two properties, combined with the properties that have already been attributed to SToCC, yield a theory that shares some features with both prototype theory and theory theory about concepts; however, section 6.11 below will explain how SToCC is different.

6.7 - Conceptual Enslavement

Implicit within the notion of a concept as a structured entity - namely, an embedded manifold - is the idea that such an embedded manifold might have an internal hierarchy, i.e. that some components or aspects of the concept might be more prominent than others, in the sense that they play a more important role, do more of the work in actually constituting whatever it is the concept means or does.

What I call a *conceptual enslaver* is the most prominent component or component-cluster of a concept's internal hierarchy, and is, as such, an expression of the way a concept is most often used by a particular person. As such, an enslaver is often an idiosyncratic expression of habitual concept use. *Conceptual enslavement* is the extraction of a (contextually and/or causally important) element (or set of elements) from the narrative(s)

provided in experience, and establishing it as the highest-ranking member in the internal hierarchy of a concept.

As such, enslavers are often derived from actually perceived objects or events, the best example available to a subject in his experience[NOTE 48], which can fulfill the function of a paradigm case, around which a concept might be built up. Eventually, the concept might be refined to such an extent that an abstract, prototype-like locus in conceptual space might come to represent more accurately what the concept is supposed to denote, but the initial 'best perceived example' will serve as the historical impetus of the concept, and at any given time there will be such a example (or class of them) which will reside at the top of the kinds of lists people give if they are asked to provide good, concrete examples of what a particular concept is supposed to mean.

Using an enslaver as the defining core of a concept allows the provisional use of a concept by referring to such best examples (e.g. characterising or classifying birds on the basis of the sparrows you see in your back yard every day), which can be subject to revisions, expansions, detail shifts or even expansive redefinitions, if contextually demanded. When an enslaver grows to be more established, it tends to function like an 'essence', to mask the *absence* of a detailed definition, or exhaustive list of typical features[NOTE 49]. This is the reason why SToCC does not fall prey to the 'missing prototypes'-problem, that does plague Prototype Theory (see section 2.3, and 6.11.2 for a closer look at this issue).

For instance, the meaning of the highly elusive concept 'justice' might, for a particular person's everyday use, be explained (or justified) in terms of a nonspecific moral stance somehow distilled from parental or religious guidance, or even the morality displayed by fictional characters. That way the enslaver of 'justice' can be a set of (privatised, possibly idiosyncratically modified) experiences and examples that might contain such disparate elements as the 'I have a dream'-speech by Dr. Martin Luther King, Jr., the particularly memorable punishment I received for breaking a window when I was six years old, a report on the evening news about the conviction of a murderer, perhaps even the climactic scene from a movie where the hero, in an act of compassion, spares the villain's life. These salient images and memories, derived from the agent's experience (aspects of a *narrative* - see section 6.7) are constitutive of a more or less vague idea of what amounts to 'justice' (or at least 'just behaviour'), hence form a comparison class based upon which behaviour (that of yourself or of another) is judged to be just or not. And when someone is asked to specify what their use of the concept 'justice' amounts to, this class (i.e. the conceptual enslaver) is used as a basis for the inferential process of justifying the use of this particular concept: examples from the concept's jurisprudence might be provided by way of explanation.

The reconstruction of a narrative by way of justification of the use of a particular concept (or set thereof) in a particular way involves expanding

upon a conceptual enslaver cache, and this reconstruction is a *social* process, i.e. it is only one pole of a process of interactional co-construction in which, usually, multiple agents are involved.

Being a low-detail placeholder, an enslaver can explain how someone can be said to hold a concept without explicitly holding the concept's complete and correct associated theory (e.g. what an experienced judge would provide as explanation of the concept 'justice'), which is the case infinitely more often than that we use a concept and are immediately aware of all it contains and implies.

Another interesting property of an enslaver is that it can act as a kind of anchour, helping a concept or subconcept resist modification of its meaning in everyday use even when the inferred accounts change. For instance, nowadays the bubonic plague is understood as a disease that is caused by the enterobacteria *Yersinia Pestis*, whereas historic outbreaks of the disease (e.g. the Black Death, which killed about a quarter of Europe's population between 1347 and 1350) were, at the time, judged to be the effects of divine retribution[NOTE 50]. So, the connected *inferred account*, hence an important part of what the concept 'plague' is understood to mean, is different now than it was six or seven centuries ago, but there is a case to be made for the idea that someone from the Middle Ages and someone from today can both use the concept 'plague', and both *mean the same thing*. This is because the enslaver for this concept is likely to involve particularly salient elements, such as the catalog of symptoms associated with the plague, that *both* people can use to indicate what they mean (e.g. 'plague' means fever, swollen lymph nodes and so on), despite the differences in inferred accounts (bacteria vs. deity)[NOTE 51]. At least part of the reason of this type of similarity in concept possession (that is, despite non-trivial differences in inferred accounts), has to do with the level of detail at which a concept is used in the majority of everyday situations. I call this a concept's granularity - the topic of the next section.

In this description, enslavers appear to be similar to the prototypes from Prototype Theory (described in section 2.3). They are indeed similar, but there are also some important differences. The main differences of these 'enslavement'-properties of concepts in SToCC with the 'prototypes' from Prototype theory about concepts will be explained in section 6.11.2 below.

6.8 - Granularity

The ability to conceptualize the world at different levels of detail - different *granularities* - and to pick the appropriate one in a given situation, is an essential feature of our intelligence (Hobbs, 1985). This feature is also highly relevant to the meaning and use of concepts, because a significant aspect of the semantic content of a proposition can shift in aspectual structure due to such granularity-shifts or -transitions. Compare:

'(1) John was sick.
(2) The virus attacked John's throat, which became inflamed, resulting in laryngitis, until the immune
system succeeded in destroying the infection.' (Croft 1991; page 163[NOTE 52])

These sentences describe the same 'thing', but at different levels of granularity, and - this is an important point to be taken note of here - as concerning different kinds of events. (1) is a low-detail description of a *state of affairs*, whereas (2) exhibits the aforementioned shift in aspectual structure by providing a higher-detail description of a *series of processes*.

In other words, the differences between using either (1) or (2) extend beyond merely being about the same 'thing' at different size levels. If I inquire after the condition of my friend John, because I knew he wasn't feeling well, I might get an answer like (1), which tells me something about John *as a person*. If I am a doctor, asking about John not (just) as a person but as a composite of organs that needs to be cured of its ailment, I might be told something like (2), which tells me something about the functioning of John's proper parts. The shift in levels as such concerns different *aspects* (these might be either aspects of the whole, or part-aspects) of the same 'thing', and a description of one aspect (say, 'John' as a irreducible primitive, namely a person) does not necessarily translate to the other without the loss of important contextual information.

Recall that I introduced the term 'contextual primitive' in section 6.6 - there I said that "(...) it is possible to specify a particular surface spectral reflectance profile, and have this be a perfectly acceptable explanation of why an object looks to have a specific colour, even though the very concept 'surface spectral reflectance' implies many more theoretical notions and relations". That is what 'granularity' is: a conceptualization of some object, process or state of affairs at a level of detail that befits the situation, and the agent's body of knowledge pertaining to the entity in question.

That means that it should be possible to define a *granularity gradient*, a trajectory of increasingly detailed conceptualizations, in some cases possibly continuous but in most cases probably discrete in character, which corresponds to the exploration of a subsection of conceptual space. It is important to understand that exploring the 'hierarchy' of ever more detailed definitions constitutive of a concept's inferred account is not necessarily the same as moving towards the non- or low-conceptual (perceptual/phenomenal) basis of conceptual space. Rather, 'digging deeper' along the granularity gradient means 'zooming in' on some subsection of conceptual space, and exploring the finer-grained structure of a (sub-)concept. A helpful visualization might be to think of conceptual space like a capillary system (the lungs or blood vessels, for instance), with the nonconceptual, perceptual spaces forming the 'trunk', and traversing along the granularity gradient as successive, increasingly detailed views of a finely veined subsection of the space.

The structure along which the zooming-in can occur consists mostly of implicatory, inferential, definitional connections. These inferential connections might get rather complex, especially if they are non-local - that is, defined in terms of other kinds of objects, entities or relations. Part of the inferred account of the (non-complex) concept 'tiger', for instance, involves reference to similar-styled and intuitively implied concepts like 'cat' and 'furry animal', but may also involve rather different kinds of concepts like 'DNA', 'reproduction', and inferences involving evolutionary history, preferred niche, animal rights, and so on, depending on how detailed and/or broad the inferred account gets. Each of these concepts has its own inferred account, also possibly with trans-categorical (i.e. non-local) inferences; this results in a highly tangled conceptual space structure.

Some of these inferential transitions in the structure of conceptual space might be of a kind that implies the concepts involved resist reduction[NOTE 53]. Conceptual superposition is a special case, involving a relation of a similar kind: a superposed concept caps (a particular subsection of) the granularity gradient at a low-grain end, but there is a kind of discontinuity between it and the rest of the gradient: the relation between the superposed concept and its subconcepts *is* inferential, according to the properties explained in section 6.6, but not straightforwardly inclusive.

In a somewhat stricter formulation, granularity concerns a mapping between conceptual space and the particular contextual demands that the environment places on occurrent conceptual dispositions. Seeking out a specific location along the granularity gradient involves relating conceptual space to the agent's context of action, to which the concept at that particular granularity is most suited. For now, this description should suffice; in chapter 8, the introduction of the 'Radicality Manifold'-model will yield a more concrete characterisation of the mappings and relations that connect conceptual space with the agent's environment.

There is one final property of granularity, in conjunction with conceptual enslavement, that I will highlight here: now, it is possible to clarify how someone's use of a complex concept need not follow the guidelines pertaining to complex concepts as set out above. If a person's enslaver of, say, the concept 'colour' is of a low enough granularity, and he is never (or has not yet been) placed in a situation where the use of a finer-grained, hence contextualised subconcept is required, he will not view the concept 'colour' as a complex concept. Similarly, it is not difficult to imagine a case in which someone is never in need of contemplating the differences between 'time' as a variable in physics, a substantivalist theory of time as an interpretation of the General theory of relativity, and the knotted topology[NOTE 54] of 'lived', phenomenal, remembered time; this person would not think that 'time' was a complex concept.

The main differences of these 'granularity'-properties of concepts on SToCC with the 'basic level categories' from Prototype theory about concepts will be explained in section 6.11.2 below. The properties of (complex) concepts

as described so far suggest particular ways in which concepts and conceptual space expand and evolve; the following few sections are devoted to various aspects of concept development.

6.9 - Concept Development Part 1: From Sensorimotor Acuity to Conceptual System

The discussion of 'enslavers' and 'granularity' above concerned the properties and dynamics of evolved and functional conceptual spaces. Obviously, there also needs to be a story about how such conceptual spaces came to be the way they are. Recall that in section 6.4, I suggested that that even higher-order concepts are ultimately rooted in basic sensorimotor contingencies. One way of supporting that claim is by providing a plausible account of the way in which, throughout a child's development, more advanced concepts can emerge from fairly basic sensorimotor contingencies. Based on suggestions by Mandler (2007) and incorporating the SToCC vernacular developed in the preceding sections, I will now present a brief overview of exactly that: the way in which an advanced conceptual system might emerge from humble sensorimotor beginnings. That is: right now I will focus on the development of a conceptual system *as such*, whereas after that, section 6.10 will offer a description of the refinement of concepts in terms of a progressive segmentation of conceptual space, inspired by Jameson's IDM (see section 5.2).

The *sensorimotor apprehension of motion* forms the foundation upon which several mechanisms for the expansion and refinement of conceptual space are built. In order of discussion, the components are:

[Stage 1]: sensorimotor apprehension of motion
[Stage 2]: correlation of sensorimotor knowledge and linguistic encoding
[Stage 3]: embodied and embedded crossmodal mapping
[Stage 4]: correlation of embodiment and abstraction

[*Concept development Stage 1: sensorimotor apprehension of motion*] Mandler (2007) states that in the earliest stages of their development, children tend to be less focused on the details of what objects look like (despite having the perceptual capacity to do so), but appear more focused on what those objects are *doing*: motion, *novel* motion in particular, attracts attention above everything else. At two months old, many children pay closer attention to objects that move in ways independent from their own motion, and at three months, they can already distinguish between biological and non-biological motion. At six months, they can pay attention to the beginning and end of an object's trajectory, and at nine months, many children are surprised when a non-biological object starts moving on its own. Mandler (2007) uses this developmental sequence to claim that a relatively simple tendency to attend to motion can help generate a lot of knowledge, and that the first practical conceptual fission is one which distinguishes animals (with their characteristic movement patterns) from

non-animals. A progressive interest in salient features of the objects in question can help specify the content of the associated concepts: in addition to a characteristic way of moving, animals also tend to have eyes, mouths, and so on.

Apart from helping the child in getting a start at categorising the objects he encounters, his early preference for moving things also provides him with the earliest *relational* concepts, says Mandler: consider spatial notions such as 'containment' and 'attachment', but also 'physical cause', which might be derived from witnessing objects bumping into eachother.

Even an apprehension of the object-concept as such (including ideas about the persistent existence and continued motion of temporarily occluded objects: children will expect an object that moves behind a larger object to reappear on the other side) can be motion-based.

Hence, the spatial primitives that lie at the very foundation of concept formation as a strategy to interpret events, according to Mandler, are 'path' (including 'start-of-path' and 'end-of-path'), 'into-container' and 'out-of-container', 'onto-surface' and 'off-of-surface', 'up' and 'down', 'linked paths' (pertaining to objects interacting), 'blocked path' and 'motion transfer'. Hence, a good hypothesis could be that the foundation for the development of conceptual ability is formed by a sensorimotor apprehension of motion. In that vein, Gallese and Lakoff (2005) provide a suggestion on how we can conceive of basic action concepts as rooted in the sensorimotor system's activity

They claim that both abstract concepts and more concrete action-centered concepts depend on action-related neural activity for their realisation. Specifically, they state the sensorimotor system exhibits the kind of structure and activity[NOTE 55] to characterise these concepts in an appropriate way.

Gallese and Lakoff discern the traditional idea of 'concept' from 'schema'. The former they describe as an internal (disembodied) representation of something external, whereas a schema is inherently interactional, in a way that depends on the way our bodies and our brains are put together, and how we interact with the world, both physically and socially. These schemata can do the work concepts (used to) do, and the parameters and their values used to define these schemata can be linked to the *functional structure* of the sensorimotor system's activity.

It should be obvious that the definition of schemata given by Gallese and Lakoff appears to be quite compatible with the definition of concepts used in SToCC, at least in the sense that concepts as I understand them are also defined in embodied and embedded terms; the definition of 'concept' they argue against is a disembodied representation of the old cognitivist kind. For the purposes of this book, their use of 'schema' vs. my use of 'concept' is merely terminological.

Gallese and Lakoff's schemata are depictions of the functional roles of relevant neural clusters, and these clusters are characterised by parameters and their values.

For example, they decompose the 'grasp'-concept/schema as follows:

"*The grasp schema.*
1. The role parameters: agent, object, object location, and the action itself.
2. The phase parameters: initial condition, starting phase, central phase, purpose condition,
ending phase, final state.
3. The manner parameter.
4. The parameter values (and constraints on them)." (Gallese and Lakoff 2005)

And:

"The various parameters can be described as follows.
Agent: An individual.
Object: A physical entity with parameters: size, shape, mass, degree of fragility, and so on.
Initial condition:: Object Location: Within peri-personal space.
Starting phase:: Reaching, with direction: Toward object location; opening effector.
Central phase:: Closing effector, with force: A function of fragility and mass.
Purpose condition:: Effector encloses object, with manner (a grip determined by parameter values and situational conditions).
Final state:: Agent in-control-of object." (Gallese and Lakoff 2005)

Below, under 'Concept development Stage 4' (correlation of embodiment and abstraction), this idea will be expanded to apply to abstract concepts.

[*Concept development Stage 2: correlation of sensorimotor knowledge and linguistic encoding*] An important step towards expansion and refinement of this basic mode of agent-environment interaction is the acquisition of linguistic skills, which can force the child to start paying closer attention to the details, rather than merely the broad motion-based properties of observed objects. After all, notes Mandler (2007), a one-year-old child might call many kinds of self-moving, interacting entities 'doggie', thus placing dogs, cats, guinea pigs and rabbits in the same category. The child's parents, rather, *do* differentiate between these various animals, using a different linguistic label for each. These different labels, and the additional features that belong to each of the various tokens of the child's primitive 'doggie'-type which parents might draw attention to (a dog says 'woof', a cat says 'meow', and a rabbit says very little but has long ears and a bushy tail), will serve to draw the child's attention to the fact that these tokens do indeed differ, and might require their own concept. Each of these concepts

('cat', 'dog', 'rabbit' and so on) having its own linguistic label will help solidify and stabilise these concepts, especially when the child notices a reliable correlation between his linguistic utterances and approval-expressing reactions of the parents ('yes, very good, that *is* a cat!'). Increasing knowledge of each concept's associated object features will constrain the generalisations made by children: at some point the child will realise that certain things which might be true for dogs can no longer be held to apply to cats (for instance 'retrieves balls that are thrown away', or 'likes to swim in murky ponds').

Mandler describes another form of language-based conceptual expansion and refinement: the development of linguistic skills will aid in connecting mainly phenomenally specified concepts to their linguistic labels. He says that learning colour words, for instance, will establish indexical relations between colour concepts on the one hand, and otherwise unanalysed colour qualia on the other. I would like to suggest an adaptation of this particular way of concept development that is less representationalist, focusing not on indexical relationships as such, but on qualia-related behavioural patterns that are appropriate to the contexts in which they are usually applied. That is: as discussed in chapter 4, colour concepts, which develop relatively late in a child's development, encode how colour qualia are linked to the objects that are perceived to be coloured by way of a behavioural prescription, specifying the appropriate patterns of action-based agent-environment interaction if confronted by said coloured object (examples being: having the ability to pick a red piece of fruit instead of a green one because you know 'red' means 'ripe/edible', or knowing how to make the correct kinds of theoretical inference pertaining to 'colour'). Acquiring the linguistic labels then stimulates the progressive refinement of conceptual (i.e. perceptuo-behavioural) space as it pertains to colour. Because, on my account, the link between qualia and linguistic labels is not an indexical relationship (colour words do not refer to colour feels alone), but the expression of a behavioural interaction practice, I would not count this as a separate mechanism for concept development. Rather, this is a special case of the mechanism described above, i.e. the stimulating impetus on concept development of the acquisition of a linguistic encoding scheme.

[*Concept development Stage 3: embodied and embedded crossmodal mapping*] Crossmodal mapping is a major source of new, more complex, higher-order, and possibly even abstract concepts. This is, to a large extent, the point made by Gallese and Lakoff (2005), already referenced briefly above. Gallese and Lakoff, extrapolating work done in Lakoff and Johnson (1980; 1999), claim that higher-order concepts can be interpreted as utilising metaphorical transformations linking them to lower-order concepts, i.e. the ones informed by somatosensory and/or sensorimotor processes:

"The sensory-motor system can characterise action concepts and, in simulation, characterise conceptual inferences. And the concepts characterised in the sensory-motor system are of the right form to

characterise the source domains of conceptual metaphors." (Gallese and Lakoff 2005)

The cognitive task achieved in using conceptual metaphors (i.e. cross-domain mappings, e.g. the description of an abstract process with an action metaphor, like 'The Eurozone *fell* into a recession') might be related to the synaesthetic confluence of sensory information in the modality-integrating function of consciousness, with a special role in that integrative process reserved for the sensorimotor system[NOTE 56].

Another interesting parallel to be drawn is with Jameson's cross-dimensional mapping hypothesis. In section 5.2 I discussed that Kimberly Jameson suggests tetrachromats in a trichromatic world use a cognitive four-to-three-dimensional mapping function. In my discussion of that idea, it was claimed that the possibility of tetrachromatic humans which would, for Jameson, necessitate such a strategy should not (currently) be overstated. However, for an understanding of shifts in granularity, or even the reduction of a conceptual superposition to the temporary, contextual usage of a particular conception, Jameson's idea might be illuminating.

Recall that the general idea, which Jameson develops in her (2005), is as follows. Dichromats and trichromats differ in the sense that individuals from the latter category are able to make distinctions that individuals from the former category cannot. That is, colour samples that, for a dichromat, lie within the same discrimination tolerance (the region in perceptual space of matching colour samples), can lie in separate regions for a trichromat, and this is due to the fact that the trichromat has an additional chromatic sensor type on the retina, because of which the trichromat's perceptual space is more complex, enabling him to make finer-grained colour distinctions.

Now suppose the majority of humans were dichromats (of a specific type), then colour language would be 'dichromatic' as well: only the colours these dichromats could distinguish would be encoded linguistically. A trichromat would have no problems using this dichromatic language, for all the colour discriminations the dichromats can make, he can make as well. He might be confused about the fact that some colours he can distinguish as being rather different are nonetheless grouped together and given the same name, but achieving this translation from 'trichromatic colour experience' to 'dichromatic colour language' is a trick that should be relatively easy to learn.

Jameson then claims that tetrachromats (living in a trichromatic society), who have a more sophisticated colour perceptual space than normal trichromats, would need to achieve a similar reduction of 'tetrachromatic colour experience', to 'trichromatic colour language', grouping certain distinguishable colours together just because there are no words to cut colour space as finely as the tetrachromat's experience does.

For granularity-shifts in conceptual space, something similar might happen, but because these shifts and cross-dimensional-mappings occur in cognitive, theoretical terms rather than in terms of innately specified perceptual dimensionalities, they are much more frequent than any of Jameson's scenarios would be. In terms of the theoretical tools of SToCC that we have discussed so far, this would involve the mapping of 'high-dimensional' or fine-grained talk in terms of a conception that encodes detailed theoretical knowledge and hypotheses about a subregion of the colour perception process onto the 'low-dimensional' or coarse-grained talk in terms of the much less specific and complex superposed colour concept[NOTE 57].

This involves the fact that, when speaking about colour in general, the language we use does not cut nearly as finely as the separate theories about various aspects of the colour-perception-dynamic would prescribe: a shift from a conception to the superposed concept is one of decreasing detail, and involves ignoring knowledge about colour encapsulated in the various subdomains.

The translations in conceptual space are more complex than the ones in perceptual space, because it is also possible to traverse from one concept to the other – the main link between the two, SToCC would claim, is due to the fact that they are both contributors to the superposed concept.

Now, crossmodal mapping conflicts such as the ones indicated above can occur when a more detailed theory (or more or less coherent explanatory account) about some phenomenon is devised, necessitating the 'invention' of new (sub-)concepts - this is one means of concept development.

[*Concept development Stage 4: correlation of embodiment and abstraction*] An additional refinement of conceptual content[NOTE 58] can derive from the extrapolated interpretation and implementation of phenomenal awareness, as it operates in agent-environment interaction. Mandler asks us to consider an abstract concept such as physical cause. His suggestion is that a deeper apprehension of this concept can result from the amalgamation of, on the one hand, the child's basic, sensorimotor intuitions about the difference between autonomous motion and caused motion (see above, under stage 1), and, on the other hand, the child's own phenomenal experience of bumping into other objects. A conceptual understanding of the physical causation involved in seeing one object causing another object to start moving can then be ameliorated by the 'oomph'-like feeling remembered from earlier encounters with walls and table legs, and/or the work required to overcome the inertia of movable objects such as toys.

This correlation of embodiment and abstraction can take the form of an embedded manifold, in a schema that can be extrapolated from the suggestions done by Gallese and Lakoff (2005) about the way in which more complex concepts are based on action-based sensorimotor activity. This is how that might work for colour-related behaviour. The colour

perception/cognition/action-process will, in most cases, be a multi-stage process consisting of various action-processes (perceiving, thinking, deciding, reaching, grasping), which might each have its own schema. Hence, the schema for a colour-perception/cognition/action-process will itself be an *embedded manifold* (see section 6.6) containing various sub-schemata. However, it is possible to simplify and summarise two reasonably basic scenarios involving colour-perception as follows:

Scenario [I]: obtaining a desirably-coloured object, e.g. a piece of fruit.
Scenario [II]: fleeing from a 'non-desirably-coloured' object, e.g. a predator.

Agent: An individual.
Object: A physical entity with parameters: shape, size, colour, visual texture, desirability. 'Desirability' is a higher-order function of the way in which the agent relates to the object, with, in these cases, either a positive [I] or a negative [II] value. This value depends on an apprehension of the goal-state of the associated scenario, which could, for instance, be obtained by way of a simulation - imagined re-enactment - of the process.
Initial condition:: Object Location: Within peripersonal space. In this case, this is larger than tactile peripersonal space, for it is determined by visibility and occurrently relevant distance, i.e. whether the individual can reach the fruit in a sufficiently easy manner [I], or whether the predator is close enough to be a threat [II]).
Starting phase:: Deliberated, directed movement: Towards [I] or away from [II] object location.
Central phase:: Executing grasp schema [I].
Purpose condition:: [I]: Apprehension of desired object: Effector encloses object, with manner (a grip determined by parameter values and situational conditions); [II]: reaching a safe(r) location
Final state:: Agent in state of relative well-being

Note how, for these two somewhat more complex action scenarios, various action-, perception- and cognition-depicting schemata (locomotion, reaching and grasping, goal-state assessment, mental simulation) are embedded within the overarching schema. The increase in complexity/dimensionality needed to model these action patterns in conceptual space would require reference not just to colour, but also to object shape, the visual context of the object, and more in general the affordances the object in its environment represents to a particular agent.

A deeper exploration of the way in which higher-level concepts relate to the pre-conceptual level of sensorimotor activity is a highly complex affair, but I hope this discussion of the Gallese/Lakoff model functions as a kind of intuition pump, that causes the idea of a conceptual spectrum (and its development) to make sense. It should be noted that for SToCC, a reduction all the way down to the basic level is not needed for definitional purposes, because in SToCC, definitions are given at a contextually determined granularity, including a justificatory structure (see section 6.6).

I can currently make one brief additional suggestion. Adapting the 'conceptual metaphor'-idea mentioned above (Lakoff and Johnson, 1980, 1999; Gallese and Lakoff, 2005) I can offer a *rather tentative* hypothesis about the *evolution* of higher cognition. The idea is that neural activity for objects within an agent's reach, i.e. within peripersonal space, is markedly different from activity related to objects outside that reach. Providing a test subject with a tool, e.g. a stick, will change the contours of peripersonal space to include the extended reach provided by the tool (Maravita and Iriki, 2004). This is thought to result in a modification of the way in which space and spatial relations are 'represented' in the agent's brain. Furthermore, an agent's *canonical neurons* exhibit specific activation patterns when objects are present that afford action (Gallese, Fadiga, Fogassi and Rizzolatti, 1996). My idea, which might be interesting to attempt to corroborate empirically, involves whether these neuronal activation patterns of action of manipulable objects within peripersonal space, as implementations of Gallese and Lakoff's (2005) 'grasp' schema, have anything in common with the metaphorical extension of *cognitive* 'grasping'. That is, perhaps there is a structural similarity and/or evolutionary continuity between the neural activation patterns associated with the following successive stages: (1) manipulating an object within peripersonal space; (2) observing a physically manipulable object within peripersonal space; (3) observing a *cognitively comprehensible* object *outside* peripersonal space.

Apart from the SToCC-specific version of Lakoff and Johnson's modal metaphor-based idea that is described above, there are several other ways in which SToCC can accomodate the creation of more complex and/or abstract concepts. After all, a substantial challenge for any theory of concepts is how to explain the existence of concepts of unobservables ('electron') or abstract concepts ('truth', 'justice', or mathematical concepts) - concepts which are not linked to or composed of specific perceptual representations in any obvious way.

Jesse Prinz (2002; see section 10.3) notes that we are capable of developing and using concepts without full knowledge of their referents. We accomplish this feat by attending to reliably correlated properties, an activity called 'sign tracking'. There are several different kinds of such signs. Often, we use superficial appearances to classify and conceptualise, focusing not on the properties that are necessary and sufficient for something to count as an example of what it is perceived to be, but on merely contingently but reliably connected properties. For instance: the concept 'human being' is usually specified not by reference to the properties of the human genome, but in terms of a general human-like appearance, including the presence of a certain number of limbs of a specific size, shape and functional applicability, characteristically human facial features, and so on.

We can also use perceivable correlates of inperceivable properties (someone's 'being humourous' can be detected by determining the proximity and frequency of smiling faces, sounds of laughter, etc.), or scientific instruments (in the case of electrons, quarks, weak

electromagnetic fields, quasars at the edge of the observable universe and other such entities, which cannot be detected without aid). And finally, words in natural languages can be used to keep track of complex or abstract concepts: I need not have personal experience with or intimate knowledge of certain entities, as long as there is a word available that picks out said entity, the exact meaning of which has been defined by experts.

In SToCC, an enslaver is what results from the practice of sign tracking; a collection of specific characteristic signs is often what constitutes the concept's narrative jurisprudence (see section 6.6). The meanings of concepts associated with unobservables, abstract concepts and even the concepts of observable but highly complex entities, processes or states of affairs are often *implicated* as a central tendency of a catalogue of observable signs that is remembered. A formal description of this implicatory tendency might be in the style of David Lewis' (1972) definition of functionalizability (which was discussed, briefly, in section 6.5): the trackable signs pertaining to some unobservable entity collectively specify a particular role for the *concept* of said unobservable to play in our accepted ways of speaking, thinking and acting involving the unobservable. For example: individual electrons cannot be observed (or at the very least not directly) by human beings, but because we have the output of instruments detecting electrons, scientific papers describing properties of electrons, memories of high school physics classes about electrons, and so on, most people have the concept 'electron' in the sense that they are capable of conversing and thinking about them at some level of granularity.

Summarising, this is the general sequence of mechanisms involved in the development of conceptual abilities from a sensorimotor foundation:

Foundation:
- sensorimotor apprehension of motion

Expansion and refinement:
- correlation of sensorimotor knowledge and linguistic encoding
- embodied and embedded (body- and motion-based) crossmodal mapping (analogies)
- linkage of embodiment and abstraction (extrapolated interpretation and implementation of phenomenal awareness, and the exploitation of signs)

In the next section, I will offer a description of the continued evolution and refinement of a particular concept (or class of connected concepts) inspired by the *perceptual* space-vernacular introduced earlier (in section 5.2).

6.10 - Concept Development Part 2: Conceptual Space Evolution

Part of the evolution of concepts, and more specifically the evolution of the array of concepts an agent might hold, can be described in terms of the progressive segmentation of conceptual space as new concepts are implemented. A suggestion for the mechanism involved in such a process was developed throughout chapter 5, in the discussion of Jameson's IDM.

Recall also that for complex concepts such as 'colour', SToCC includes the thesis that differences of categorization and definition result in an array of potentially incommensurable stories about a particular phenomenon, and that due to that incommensurability the concept of that phenomenon cannot be held to apply in full in all situations it is supposed to.

Maund (1981, 1995) draws attention to this peculiar property of the colour concept by stating that there is no single property (physical, mental, or whatever) that plays all the roles customarily attributed to 'colour'. His solution to this problem was to invoke the idea of 'conceptual splitting': the process by which a unitary, naive colour concept that was prevalent in the (distant) past split into two incompatible subconcepts at some point during our cultural and scientific development. 'Colour' actually has a physical and a psychological component, he says in his (1981) article, each with its own content and domain of application.

Like Thompson (see chapter 4.5), Maund posits the current situation in the philosophy of colour involves a particularly persistent dichotomy of objectivism versus subjectivism. To overcome the deadlock, Maund suggests we should adopt a pluralist view of colour.

He says:

"(...) there are no properties (or no good reason to believe that there are) that satisfy all of the following constraints (nor even satisfy any two of them):
1. which play the causal role required to be played by colours in the perception of colour, i.e. in the production of colour experiences and
2. which collectively have the kind of structure embedded in colour ordering systems; and
3. which have the sensuous character of colour." (Maund 1995)

And:

"(...) there is not one viable concept of colour but several. (...) we can distinguish between different kinds of colour, e.g. physical colour, optical colour, psychophysical colour and so on. (...) Each serves its own purposes and functions." (Maund 1995, pg. 114)

So, Maund notes that the word 'colour' has several different uses. It is easy to see how this can be the case: the colour perception dynamic is a complex process spanning different kinds of processes. As noted before (section 6.1) the descriptive and explanatory accounts to understand these processes derive from different branches of science - physics for surface spectral reflectance, neurophysiology for the neural processing of retinal stimuli, phenomenology for the ‘feels’ of colour percepts, anthropology and linguistics for the segmentation of perceptual colour space, evolutionary biology and philosophy to provide an interpretation of the meaning of it all, and so on. These different branches of science are not compatible, in the

sense that they use different kinds of research methods and tend to define their terms differently: for current purposes, it suffices to note that each of these theories attempts to say something about colour, but that they all use the concept ‘colour’ in different ways.

This boils down to the fact that, for instance, ‘colour’ defined in terms of surface spectral reflectance can not, in any obvious way, be translated into or connected with phenomenal ‘colour’. Part of this incompatibility can be explained by the classic form of metamerism as applied to colour: the vast, potentially infinite dimensionality of object reflectance space is reduced to the relatively compact dimensionality of retinal receptor space, which means that a particular perceived colour cannot be linked in any consistent fashion with a specific reflectance, or the physical structure underlying it.

However, the problem of an incompatibility of dimensionalities and space characteristics of the various ‘components’ of the colour perception dynamic is more widespread that that: there are discrepancies of that kind between almost every stage. The infinite dimensionality of reflectance space clashes with the three-dimensionality of chromatic receptor space (for normal humans), which requires some kind of transformation before it might be mapped onto the four-component chromatic opponency structure of perceptual colour space. The layout of perceptual space is determined, in part, by the complex structuredness of the physical and socio-cultural enviroment, and is linked to the equally complex structuredness of *conceptual* space.

Maund attempts to explain an incompatibility of various kinds of speaking about colour (which he, by the way, does not explicitly define in terms of diveregent dimensionalities) by suggesting that there is nothing in the world that actually possesses the property ‘colour’ in the way we habitually attribute it, hence that the normal idea of colour refers to a virtual property. What this means and why Maund’s ‘solution’ does not work will be explored further down, but for now it suffices to note that it is of paramount importance that we find a theory that can deal with the fact that there are various subconcepts that might not refer to the same property when using the term 'colour'. As noted, ‘colour’ is used in various kinds of science and everyday practice, each with their own area of application: physical, neurophysiological, phenomenological, linguistic, practical (red paint caused white paint to turn pink), emotional and aesthetic (green is soothing), and so on...

Still, it appears to be coherent, in some fashion, to speak of ‘colour’ as a single concept. Maund holds that an important reason for this perceived coherence of the various colour concepts is their conceptual ancestry: once upon a time, there was a unified concept of 'colour', but over time several aspects of our way of speaking about colour have diverged, yielding the current pluralist matrix. In Maund (1981), this process is called 'conceptual fission'; and alternative term (and the one to be used in the current text) is ‘conceptual splitting’.

This is what that means: we once might have had a singular colour concept, which underwent a series of segmentations to result in the current situation of an array of colour sub-concepts, each with its own domain of applicability, and in some instances mutually exclusive in scope, while at the same it makes sense to speak of colour as a singular concept.

The story developed in my account so far can be understood as an extrapolation of Maund's idea, for it includes the idea that the naive 'colour'-concept breaks apart if put to contextualised use, not in just a physical or psychological part, but in *several more* subconcepts. However, now I can say a bit more about how this process of conceptual splitting might transpire.

Recall once again chapter 5, in which I developed a description of a mechanism of *perceptual* space segmentation, and my suggestion is that this process can be viewed as a low-dimensional instantiation of the kind of process that can produce new, finer-grained concepts from old ones. Jameson and D'Andrade (1997) and Jameson (2005) in particular claim that irregularities in perceptual colour space facilitate its progressive compartmentalisation that turns out to line up fairly neatly with Berlin and Kay's (1969) evolutionary sequence.

As mentioned before, three-dimensional colour space is not a perfect sphere, with protrusions at places where the saturation for particular hues can be specified at much higher levels (in the case of red, for instance), and indentations at places where saturation levels are unavailable beyond relatively low values (in the case of yellow). Because of these irregularities, distances between foci are not uniform. If the most primitive lexical segmentation of colour space has been made by separating colours into dark/cool and light/warm categories, the most informative additional term that might be acquired specifies RED, which has a focus farthest away from the regions specified by the intial two categories. The fourth most informative colour word to be acquired would be either yellow or blue, followed by green, purple, pink, orange, brown and gray (as determined by distance computations carried out by Boynton and Olsen (1987)).

Adapted for the evolution of conceptual space, this idea would imply that new subdivisions of conceptual space - i.e. the emergence of new, finer grained (sub-)concepts -, are likely to occur if the interaction of agent and environment is such that behavioural, locutionary and/or cognitive patterns arise (or are evoked to emerge), that need to be given a structurally appropriate place in the agent's repertoire of dispositions. In brief, each iteration of conceptual space segmentation occurs in such a way to achieve a maximum increase in cognitive and/or perceptuo-motor acuity. The 'evolutionary pressure' exerted then concerns whether the formation of a specific concept will help the subject perceive/act in a more efficient manner.

Taking a cue from Vantage Theory[NOTE 59] (MacLaury, 2002), it might be helpful to say that this process of categorization is driven by the need to find a niche-appropriate balance between the opposing forces of similarity-seeking versus differentiating judgments. Similarity-seeking judgments result in fewer distinctions being made, hence coarser-grained concepts, which would benefit the expedience and cost-effectivity of cognitive processing. Differentiating judgments, on the other hand, consist in a context-driven exploration of the granularity-gradient, pressing for finer-grained distinctions in cases where detailed judgments are required[NOTE 60].

The pressure towards forming new or more finely grained concepts derives from an agent perceiving a conflict between what he wants to do, or thinks he is able to, based on the concepts he has, and what his environment allows him to do successfully. If the success of his act disappoints - if there is a misalignment between intent and pay-off -, the agent might be forced to re-examine his beliefs or behavioural strategies, hence revise the implications and relations encoded in his conceptual space. This means that the evolution of conceptual space is the expression of an embodied and embedded learning curve, in which (conscious) cognition might (but need not!) play the role of catalyst: if the misalignment registers in consciousness, this could result in an impetus of increased strength towards a fitting adaptation.

Apart from the progressive segmentation of a conceptual space with more or less pre-established boundaries, aspects of conceptual space evolution can also be described in terms of *expansion* of the space, as completely new concepts are formed, and *density increase*, as a particular concept is defined in a more rigorous and detailed fashion. All these processes can be described in terms of processes involving *embedded manifolds*.

Recall , from the discussion in section 6.4 and 6.5, that conceptual space is to be thought of as an embedded manifold, which in this case means that it is an abstract space-like structure containing smaller structural elements. So, subsections of conceptual space are themselves embedded manifolds, constituting a concept or subconcept at some granularity. It is embedded because it *co-constitutes*, with other manifolds (subconcepts), a concept at a lower granularity, and *contains* multiple manifolds, each associated with a subconcept at a higher granularity. It is the diversification, refinement and expansion of embedded manifolds at some granularity that drives the segmentation of conceptual space.

A possible cause of this diversification, refinement and expansion of conceptual space can be the increase of the space's dimensionality, as new concepts are added.This occurs, for instance, when something changes in the environment in such a way that a novel strategy for dealing with that new situation is required. So, the advent of a new and/or better theory or idea about some aspect of the world might consist in the refinement of some components of a particular concept or the structure of the manifold. Such a change could also help unlock new ways of seeing, understanding

and/or conceptualising elsewhere in conceptual space, for instance by offering new data (due to better instruments and shifts in ideas in those other areas) or new interpretations of existing data.

For example, suppose that fossils of a previously unknown species of hominid are discovered, carrying some hitherto unforeseen implications about the evolutionary history of *homo sapiens*. This means that some aspect of the concept 'human', namely the fairly high-grained aspect that says something about the species' evolutionary heritage, needs to be modified. We can suppose that at a particular grain, what was previously covered by one subconcept now needs to be described by a more complex account, involving one or more new subconcepts. That is, conceptual splitting at some granularity leads to conceptual space density increase, an increase in 'content', in that region.

This form of conceptual space evolution, i.e. involving embedded manifolds, allows more detailed descriptions, for instance of the following kind: the evolution of either singular subconcepts or semantically linked groups of them in itself, or a shift in the overarching concept, might result in embedded manifold boundary transgressions, which means that an object, event or process that was previously described by a particular concept, now falls outside the explanatory range of that concept. That is, in the new situation (either because the subconcepts' enslavers have changed or the inferred account has), some component of a concept will fall outside what we can call the *legitimization region* of the inferred acount. If the errant concept-component is still of demonstrable use in a relevant context, it might *demand* a new and fitting subconcept (plus an inferred account) to contain it. Hence, what was previously a single embedded manifold (defined by its inferred account) will now have split into two manifolds, with separate inferred accounts to go along with them.

For instance, this abstract scenario can function as a description of a way in which the concept change involved in the case of the new ape fossil as described above could occur. Suppose that the old explanation involved several sources of evidence that suggested that a group of one kind of human ancestors traveled from one region to the other, over a specific period. Suppose that the concept of what this kind of human ancestor was and did, is modified due to the discovery of some new evidence, which suggests that this group could not have visited some distant part of the region it was previously thought to have inhabited. Still, there are indications that at least some type of hominid must have inhabited that distant part of the region, and sure enough, new excavations there uncover fossils of a new type of hominid. So, a high-grained remnant of the old concept that now falls outside the boundaries of the new concept is still demonstrated to be of sufficient use, hence demands a new concept to 'contain' it.

More in general, a change in some concept or subconcept can be described in terms of:

-embedded manifold density increase (or decrease, if a subconcept that was useful before falls into disuse, i.e. is forgotten)
-embedded manifold expansion or contraction (without change in density)
-sections of other embedded manifolds being colonised/acquisitioned, or sections being abandoned/relinquished.

...or a combination of the above.

Here is another example, one that incorporates a few of these processes: suppose that a neurophysiological theory about colour is found that is (a) much more detailed than the current theories (i.e. embedded manifold density increase), (b) uses new subconcepts (i.e. inherent manifold expansion) and (c) explains away certain aspects of the phenomenology of colour perception (i.e. acquisition of (part of) another manifold; manifold expansion by conquest).

The weight of the occurrent enslavement force (see section 6.7) of certain elements of some (sub)concept is the factor that determines the difference between mere manifold overlap (some concepts might share a few elements in a peaceful, mutually 'unthreatening' manner) and actual conquest of another manifold. A strong enslaver might 'overpower' a weaker enslaver if it comes too close, for instance when a powerful new theory offers explanations for cases that were previously thought to require their own subconcepts. These are all metaphors we can use to describe the ways in which (aspects of) concepts develop.

The above implies a kind of *conceptual holism*: first, there are webs of meaning between concepts and subconcepts, and these webs often reach across manifold boundaries; second, there are various kinds of mutual multi-level influences within classes; and finally, 'shockwaves' of manifold change may ripple throughout conceptual colour space, as changes in one concept influence the contents of several other concepts.

6.11 - Intermediate Evaluation of SToCC

The foregoing concludes the cursory description of SToCC. The most important part of assessing whether SToCC is an appropriate characterization of concepts will occur in the following section of this book, when this model is used to build a framework that is intended to provide clues about how to account for higher cognition in an embodied and embedded context. For now, however, we can take a look at the way in which SToCC deals with the 'inherent instability of the concept-concept', introduced earlier, how it differs from Prototype Theory, and how it is able to answer two arguments directed against Theory Theory about concepts (see Laurence and Margolis, 1999), which resembles SToCC in some ways. Then, I will discuss three arguments against SToCC-style theories of concepts, courtesy of Fodor (2004), and finally, I will compare SToCC to Conceptual Role Semantics.

6.11.1 - SToCC and the Instability of the 'Concept'-concept

First, I will discuss SToCC's solution for the instability-issue. The claim will be that the notions 'conceptual enslaver' and 'granularity' *in tandem* allow us to see how concepts can be real, and how the previously mentioned *inherent instability of the 'concept'-concept* might be solved.

In SToCC, having a concept crucially involves *doing* something, or *being able* to do something: it involves acting or speaking or thinking, or being disposed towards such activities in particular circumstances. Having the correct concept of 'traffic light' involves successfully utilising the light signals it emits to avoid being run over by a city bus in the middle of the intersection; having an appropriate concept of 'mold' involves being able to correctly identify it amongst a bevy of other furry purplish-green substances.

In short, having a concept and possessing some grasp of its behavioural and/or cognitive implications are necessarily intertwined. Obviously, this allows for gradations in the appropriateness with which concept-possession can be ascribed to some agent: someone can be more, or less adept at understanding the implications inherent to a particular object, process or state of affairs. An ornithologist will obviously have a much better grasp than I do of what a nightingale is and which traits belong to the creature, but that *does not* mean that he, with his sophisticated bird-related concepts, and I, with my limited knowledge of them, could not be said to be using *the same concept*. The difference between the two of us to take note of here would be the potential of exploring the properties and implications of the concept along its granularity gradient, which for the ornithologist extends much deeper into the details than for me. This can be a criterium for possession of the same concept: the main concept we both use in everyday parlance would be the same if we were able to communicate successfully while using said concept, within a particular granularity-bandwidth.

Even if, due to ignorance, I were to ascribe certain properties to a nightingale incorrectly, that would not necessarily disqualify the expert and I from using what, in some relevant practical sense, could still be called the *same* concept: we would both use the same word to denote the same creature, for instance by adhering to the same conceptual enslaver (e.g. a characterisation that could be something like 'a flying animal of this size and these colours, producing such-and-such bird songs (...)'), albeit with slightly different property-ascriptions. A lot of coarse-grained practical discourse would still be possible without these differences in finer-grained beliefs having any discernable influence. However, obviously the stretchability of the concept knows bounds: when the discussion turns to the nitty-gritty of nightingale-properties, the inaccuracy of my beliefs might be exposed. In such a case, the notion that we possess the same concept is tested, and the ability to either reach a middle ground or yield to the more knowledgeable of the two discussion partners determines whether this notion of shared concept-possession can be maintained.

The message to take away here is that concepts, in SToCC, are fundamentally defined in terms of *socially shared* behaviour and dispositions towards behaviour, and that this has the important implication that the possession conditions of concepts are also defined in terms of a *communal heuristic*: two people can come to believe that they possess the same concept of some object, process or state of affairs if they can exhibit behaviour (action, locution or cognition) involving this entity, and can find no glaring or debilitating incompatibilities in their respective approaches to the entity in question. This kind of practically forged belief that there is a shared concept often suffices for interpersonal concept identification. And even if recognizing a shared directedness at some object in the way described is not enough for a judgment of concept-identity, an additional uniting, generalizing force is likely to be found in *language* (as mentioned a few times before). We like to label concepts with linguistic terms, and the finite nature of the array of available terms necessarily renders this labelling practice an act of imperfect categorization: language *does not* 'cut nature at its joints', to borrow a poignant phrase from Plato[NOTE 61]. This compartmentalization of (knowledge about) the world contains an inherent generalizing tendency, an inclination to ignore differences if they are small enough. These behavioural and linguistic tendencies together alleviate, to some extent, the inherent instability of the 'concept'-concept.

6.11.2 - SToCC vs. Prototype Theory

Back in section 2.3, the Prototype Theory of concepts (PT; e.g Rosch 1978) was explained as containing the idea that some things may be better examples of a particular category than others. An object will be counted as an exemplar of a particular concept, not just in case it possesses all the necessary features included in the concept, but also if it possesses a sufficient number of them, and sufficiently many important ones.

PT's main strength is its elegant account of categorization, which utilises fuzzy set theory to characterize similarity judgments of a category representation and an exemplar representation. SToCC bears some resemblance to PT about concepts; the notion 'enslaver' in particular appears similar to the notion 'prototype'. However, there are critical differences as well. Two general difficulties for PT were discussed in section 2.3; two additional problems are especially relevant when discussing the similarities and differences of SToCC and PT, and it is to these differences that I will now turn.

One of such differences involves the probem (for PT) that for a large number of concepts, test subjects fail to isolate any typicalities. Fodor (1981) provides the example that even though there might be prototypical cities, or even prototypical American cities, there are no prototypical 'American cities situated on the East Coast just a little south of Tennessee'. Uninstantiated ('31st century invention') or overly heterogeneous concepts ('object that weight more than a gram') fail for obvious reasons (one wouldn't know where to begin in assessing the typicality of examples), but

even more commonplace and rather important concepts such as 'belief' and 'justice' fail to exhibit prototype structure.

Enslavers differs from prototypes *exactly* because of such 'missing prototypes' cases: an enslaver tends to function like an 'essence', as explained in section 6.7 above[NOTE 62]. Recall that in section 6.7, the case of 'justice' was discussed. Now, Prototype Theory will have difficulties isolating any most important exemplar belonging to such an abstract concept, hence in this case it will not be able to specify what would be this concept's prototype. SToCC's enslaver, on the other hand, is a selection of characteristic experiences and ideas which have come to shape the agent's understanding of the concept. In case of a concept as abstract as 'justice', the enslaver does not so much 'contain' examples of justice itself, but rather of 'just behaviour'. In this case certain behavioural elements that are shared amongst these examples can be used to individuate the concept 'justice' *by proxy*, i.e. by focusing on properties of an associated phenomenon (the kind of behaviour people exhibit while acting according to just principles or in a just manner) rather than justice itself. The main difference between Prototype theory and SToCC is then that prototypes collectively lock in the *definition* of a concept, whereas an enslaver enables *inference* towards ideas and behaviour that are appropriate for someone professing to have a particular concept.

Another important argument against PT involves its inability to explain how to combine graded extensions in a way that aligns with strong intuitions about concept combination (Osherson and Smith, 1981; Laurence and Margolis, 1999). It is possible to see this problem as the difficulty, in certain cases, of Prototype theory's subsidiary, fuzzy set theory, to relate the prototype of composite categories to the prototypes of the constituent categories. Problems arise with composite concepts such as 'pet fish': a prototypical fish would be, for instance, a bass or a trout - it would possess features such as 'gray', 'undomesticated', and 'medium-sized'. By way of contrast, good examples of pets would be cats and dogs, possessing features such as 'furry', 'affectionate' and 'tail-wagging'. How does a prototypical pet fish - a goldfish, say, with features such as 'small', 'brightly coloured', 'lives in a fishbowl' - relate to the features associated with its constituent categories? The features of the composite concept are not, in any clear fashion, a function of the features of its constituents.

SToCC can state that a composite concept requires a new inferred account to provide its (initial) structure, which can depend on the accounts associated with its constituent concepts, but need not be a straightforward intersection of those accounts. The inferred account about pets defines a specific subsection of the animal kingdom, populated by creatures that might be pets according to the criteria set forth in the account. Some creatures will make great pets, according to the account (say, cats and dogs), while others would be less ideal, but still fall within the specific 'pet' domain. Within that domain, one would suppose, falls a number of fish species; perhaps none of the fish are in the vicinity, within this abstract

space, of ideal pets. Obviously, the ones that are too large and ferocious (and so on with a list of traits not applicable to pets), such as great white sharks, will not even be within the 'pet' domain. The inferred accounts, hence the structures of the specified domains in conceptual space, as well as the 'solidity' of the borders between domains (whether it constitutes a fade into gray, or rather a clear cut-off point) will be context- and agent-dependent: a very valuable, rare and fragile fish might make an acceptable pet for a wealthy collector, but not for a small child. The ways in which the composite concept relates to its constituents is mediated by the ways in which the inferred accounts of the constituents relate to eachother, *plus* the way the composite concept functions in its own context.

This difference in explanatory potential between PT and SToCC suggests that even though both theories might appear somewhat similar to eachother in a first analysis, there turn out to be substantial differences as well. This is, first of all, because in SToCC, concepts are not representational entities of the classical kind, but involve abilities. The difference mentioned is also due in part to the fact that *complex* concepts require a different set of explanatory tools than what standard PT would be able to account for. Rather than a clean compositional hierarchy of elements, colour conceptual space (for instance) contains subconcepts of colour which do not refer to the same *kinds* of entities, yet are used in a kind of suspension of judgment as if they might do anyway. In the case of colour, the role of the concept ‘colour’ or any of its subconcepts is more complex than for many ‘regular’ concepts, considering their entanglement with many different forms of science (and the active summarising, theory-constituting and theory-generating roles the colour concept[-s] take on), the tension between universalist and relativist tendencies in terms of intercultural manifestation (see chapter 4), and (importantly) the strong intuition that colour involves a fact of the matter about the world rather than being merely a by-product of quasi-idiosyncratic definition (which is the case for many ‘regular’ concepts).

Hence, there is an important way in which the account of (complex) concept formation and maintenance expressed in SToCC edges away from Rosch's PT. In PT, the relations between the various strata of the conceptual mereology are fairly straightforward: they can be characterised in terms of *inclusion relations*, resulting in a taxonomic-tree-like structure in which a lot of particulars are gathered, on a higher level, under a single banner, and these classes themselves might be grouped together in some way on an even higher level. At least for colour (and other complex concepts), the suite of relations between the strata, as well as the content and internal structure of each stratum, is much more complex. A concept is not just a categorization tool, but (also, in some sense) a functional entity following (and helping to create) a specific set of application rules - this active and interactive power of concepts was characterised earlier (section 6.3) in terms of dispositions for embodied and embedded action of an agent. In other words, where in PT the internal structure of a concept is determined by generalized definitional categorization, SToCC suggests a structure based on interactively idiosyncratic explanatory inference or justification,

involving the construction of enslavers from the agent's narratively arranged action-context.

The tangled conceptual structure that results, including the contents of particular embedded manifolds, cannot be explained away merely by a blanket referral to the Theory Theory of concepts (TT); as expained in section 6.6, SToCC uses *inferred accounts* instead, and the differences with TT will be explained in the next section.

6.11.3 - SToCC vs. Theory Theory

This part of the intermediate evaluative discussion involves the way SToCC can deflect attacks on the (in some ways similar) Theory Theory (TT) about concepts. During the earlier discussion of TT, in section 2.4, I mentioned a few problems that this theory needs to face. I will now briefly mention the ways in which I feel SToCC is able to parry these attacks, before diving into a discussion of a more serious attack on TT, by Kwong (2006).

Recall that for TT, the way in which the essence placeholder picks out an extension is judged to be overly sketchy. In SToCC, an enslaver need not be sketchy at all, but still function as a 'placeholder' in the sense that it stands in for the totality of the concept, and by proxy for the inferred account. An enslaver can come to function like a meme[NOTE 63], buttressed by the most characteristic case or cases from the concept's jurisprudence, and might as such be transmitted in a form with properties that, from the point of view of the (highly detailed and carefully constructed) theory, are of secondary importance. For example, the concept 'bird' might have the enslaver 'sparrow', because for that person the characteristic examples of birds are the sparrows he sees flying around his backyard every day. Hence, the primary memetic properties he associates with bird are 'flies', 'has wings', 'lays eggs', 'dwells in trees', and so on. In everyday situations, judgments about whether or not something is a bird might indeed be based on its resemblance to the sparrows he knows so well. However, when we need to know what kind of a thing a bird *really* is, it appears a scientific judgment is warranted, so a DNA test or a consult from a ornithologist might be in order. Thus, the context of the classification judgment determines the richness that is required of the concept, but even in the most superficial tenure of the concept there is a clear path towards the underlying account.

The related problem for TT, that some people might represent incorrect information as part of their essence placeholder can also be dealt with by SToCC. After all, in SToCC there is some flexibility in the connection between a concept and its inferred account - enslaver stability can ensure survival of the concept even under shifts in the kinds of things the inferred account picks out. An important reason for this ruggedness is the explicit acceptance, in SToCC, of the historical and dynamical aspects of concept possession (via the narrative that supports the concept). An inferred account might very well be wrong (in fact, SToCC can say that the relevant kinds of accounts are, by and large, hypotheses subject to modification), but

that does not change the fact that said account was explicitly intended to explain or describe a specific state of affairs in the world: 'this story is about that phenomenon *right there*!'. The specific phenomenon a concept is supposed to be associated with will still be there after a shift in the way an agent wishes to account for said phenomenon (for instance, because he discovers a new and better theory about it), and the way the concept points towards the phenomenon in question could still be intact, so if said change stays within certain bounds, there is no inherent difficulty in allowing different accounts to be associated with the same concept, hence with the same phenomenon - all this as argued for in section 6.6.

In fact, SToCC is committed to the idea that concepts and their inferred accounts evolve - there is a continuity of practical contexts in which a specific concept is used, and a continuity of the roles said concept assumes in those contexts. As long as a concept and its inferred account, or a sufficiently large part of that constellation, remain within that utility continuum, we can continue to speak of *the same concept*. When there is a large schism - say, a significant paradigm change - perhaps the re-use of the name of the concept might offer some modicum of continuity, but then the case for the assertion that a wholly new concept has emerged can also be made.

However, in the majority of cases it should be possibe to identify some concept as the same one from instance to instance, for the reasons outlined above, despite changes in the inferred account. Similarly, some people might attribute incorrect beliefs (which would, effectively, constitute using a different explanatory account) to what they believe their concept to denote, but still hold the same concept as someone who *does* utilise the correct account (where an account's correctness is judged by, say, the overwhelming majority of experts). And even in that case it should be possible to weed out blatantly false explanatory accounts soon enough, simply because they fail to provide good explanations in the relevant contexts, and interactions with other, more knowledgeable concept-users will usually generate the appropriate feedback.

With this explanation, I believe I have also deflected the second attack on TT: the problematic stability of concepts under theory-change.

Now I turn to a more potent attack against TT, which is launched by Kwong (2006). With his arguments, he aims to support Fodor's (e.g. 2004) contention that there are currently no tenable theories about concepts. Kwong distinguishes two variants of Theory Theory about concepts: the *literal* variant states that concepts are structured analogous to scientific theories (Gopnik and Wellman, 1994), and the *liberal* variant likens concepts to theories on much looser criteria, which count the power to provide explanatory relations and the general capacity to account for conceptual correlations as a theory of some kind (Laurence and Margolis, 1999).

The thought experiment he uses to support his arguments involves James, who has acquired a fair bit of scientific knowledge about mold, and therefore qualifies as possessing the concept 'mold' even on the literal view. However, his grandmother shrieks when and only when she is in the presence of mold, which for James comes to count as a highly reliable mold-identifying clue, and the *only* mold-identifying clue when he encounters a kind of mold he is unfamiliar with. So, for James, his grandmother's shrieks come to represent an essential component of his categorization-behaviour involving mold, even though these shrieks are not in any way necessarily connected to the scientific theory.

Against the literal view, this scenario levels the objection that, in its strictest formulation, it does not allow such idiosyncratic additions that do not fit in a proper theory. Even if the criteria for what would count as a theory are given a somewhat looser interpretation to allow inclusion of these kinds of contigent, but empirically relevant beliefs, the literal view of the Theory Theory comes up short: this would result in an uncontrollable expansion of conceptual structure, violating the laws of cognitive economy.

This latter point is also the main objection against the liberal view of Theory Theory: if any and all beliefs that are relevant to a particular concept or categorization task are to be represented in a concept's internal structure, this will result in extremely cumbersome entities, which would place extreme stress on cognitive processing capacity; tests measuring processing time do not show these effects.

I contend that SToCC is capable of resisting these arguments. The easy way out would be to note that SToCC does not view concepts in general as representations of the classical, internal kind, hence does not require anything to be 'represented' under some concept, but there is room for a more substantial counter-argument. Seeing why this is so begins with understanding that the notion 'inferred account' in SToCC differs from the notion 'theory' in Theory Theory. To recap, the idea is that an inferred account is not necessarily a fully realised scientific theory (or even folk-psychological theory, if there is such a thing), but consists in the capacity of meeting appropriate standards of accountability. This is the main *general* difference of SToCC with at least what Kwong calls the *literal* interpretation of the Theory Theory of concepts. Perhaps SToCC exhibits closer resemblance to Kwong's *liberal* explication of Theory Theory, but SToCC contains a few theoretical components that allow it to accept much of the scenario of Kwong's thought experiment, and explain why it does not automatically fall prey to the weaknesses he highlights, where even the liberal variants of Theory Theory do.

These elements include, as explained, the fact that for SToCC, concepts are not theories, but capacities that are structured in conceptual space in terms of inferential relations (and yes, sometimes in terms of relations prescribed by actual theories, especially where it involves scientific

concepts). Also, it includes enslavers and granularity as structural elements of the theory.

I can substantiate my counterargument by formulating the following claim: idiosyncratic definitional elements of a concept are not necessarily problematic. One possible worry that underlies Kwong's thought experiment is that 'concept-possessing behaviour' might not be caused directly by the object/process/state of affairs the concept is of, but rather by some correlated state, where this correlation can be highly idiosyncratic. In Kwong's example, this correlation (of the presence of mold and his grandmother screaming) is reliable, and seemingly ridiculous. However, that type of a roundabout route towards identifying some object is not nearly as uncommon as one would think.

Consider the following example: even the physicists who have the most complete and accurate concept 'muon' do not possess it because of direct personal acquaintance with muons, but because they have learned to carry out experiments and interpret computer readouts in a particular way: the occurrence of some event involving muons would, if such be reliably correlated with the output of the measuring equipment that these physicists *do* have access to. The limited concept I have of muons lies at an even greater distance from the conceptualised items themselves: I have read a few articles and books about them, but I have never conducted or even witnessed an experiment that corroborated their existence. Still, for reasons given above (section 6.7), those scientists and I can still be said to possess the same concept, in some cases.

This knowledge-by-proxy is probably rather widespread, and might even be evolutionarily significant: Dooremalen (2003) argues that many complex and/or hidden but important properties are perceived by way of much more overt properties, which are sufficiently reliable indicators of the complex/hidden properties. An example of such a hidden property would be fertility, a property that is undeniably relevant to procreation. However, we (much less an animal) cannot see 'fertility' directly, so we have evolved in such a way that we react to a correlated feature, physical attractiveness. The (hormone-modulated) directedness towards this easily detected feature allows the most attractive, hence (what are likely to be) the fittest specimens of a particular population to mate with eachother, thus increasing the chance of healthy progeny. Dooremalen calls this *evolution's shorthand*: a compact and salient way of denoting a complex property that is evolutionarily significant[NOTE 64].

If this idea is correct, the kind of mediated concept acquisition Kwong uses to drive his arguments contra Theory Theory is very common. However, one aspect of his argument does remain: the idea that all knowledge that is linked or correlated to some concept is to be included in the concept's theory, hence the concept itself, results in a highly uneconomical conceptual structure. It is my claim that SToCC's enslavement hierarchy, as

well as the granularity gradient, provide enough tools to explain how a multi-denotational concept can still be encoded in a compact fashion..

The introduction of the Radicality Manifold in chapter 8 will provide a more detailed description of the connection between concepts and the agent's specific way of embodiment and embeddedness, but for now I can make my case in the following way: SToCC defines concepts in an embodied and embedded fashion. That is, in SToCC, concepts are defined in terms of an agent's capacities towards successful interaction with his environment. This means that significant aspects of conceptual structure need not be represented internally, but can be 'outsourced' to the environment, where this environment can include books, the internet, cultural customs, other people, and so on (compare Clark's [1997] 'scaffolds'). Apart from this spatial outsourcing, there is also a temporal variant, which is essential to the notion of an enslaver: having *accountability* as a central criterion of concept possession involves shunting a lot of content that is low in the enslavement hierarchy away, betting/hoping that the possibility of someone requesting an explanation (i.e. a justification of the way you use a particular concept) is not actualised. In other words, possessing and using a concept by having an economically structured enslaver stand in for a much more complex conceptual structure involves, in some sense, *bluffing* your way through everyday life, and as long as no one calls you out, there is little need to use complete representations of concepts, carrying along all those other beliefs that are implicated by and otherwise connected to your concepts. A given concept's enslaver in effect occupies the role of a more elaborate inferred account, and often the use of just that enslaver suffices for everyday use.

6.11.4 - SToCC vs. Fodor

One of the main adversaries to anyone claiming to have a theory of concepts nowadays is Jerry Fodor. In his (2004), he formulates three arguments against what he calls 'bare bones concept pragmatism' (BCP). BCP, in his description, is a collection of approaches to explaining concepts that can be characterized by the shared feature that concept possession is constituted by certain *epistemic capacities*, namely 'inferring' and 'sorting'. Whatever the differences in detail between SToCC and the various exemplars of BCP, the way the 'capacity'-angle is stressed makes it sound similar enough to SToCC to warrant a closer look at Fodor's three counter-arguments. Luckily, the preparatory work of the sections above allows me to be reasonably brief in showing how the problems described by Fodor do not arise for SToCC.

First, Fodor's 'analyticity'-argument states that if the possession conditions for a concept are defined in terms of inferentiality, neither of the two options available to characterise what kind of inference is in play (*holism* and *molecularism*) appears to be able to do the job. The 'holism'-tack would state that every inference involving some concept counts as a possession condition for that concept, but that would mean, says Fodor, that the

publicity of concepts is in danger: no one could ever be said to have the same concept as anyone else, because the chance of an exact match of the idiosyncratic set of possession conditions for concepts between two people is infinitesimally small.

SToCC accepts the (potential) idiosyncracy of concept formation and use, but provides granularity and the social practice of providing justification for concept use as properties to explain why and how two people can be claimed to possess the same concept, despite idiosyncratic differences at certain levels of granularity - see section 6.8. However, SToCC defends a kind of 'local holism' (embedded manifold change might only have local, plus selectively distal effects) that might share some features with 'molecularism', as Fodor describes it.

That is, molecularism, in this context, is a weaker claim which states that some, but not all inferences involving a particular concept can count as possession conditions for that concept; the question then becomes *which* inferences are appropriate. In SToCC, the answer to the question whether a particular inference is appropriate will be measured in terms of *achievement*: the three-tiered *achievement-criterion* provided in section 6.6, with behavioural, cognitive and phenomenal aspects, can help decide which apparently concept-based actions are appropriate, and which ones are not.

Second, Fodor's 'Compositionality'-argument claims that epistemic accounts of concept possession are incompatible with the (seemingly) non-negotiable principle of compositionality. Very sketchily, 'compositionality' involves the property of language that if we can think and speak of pets and of fish, this also allows us to think of pet fish. However, an epistemic account of concept possession is developed, in part, in terms of sorting acuity (i.e. how adept an agent is in distinguishing A's from not-A's), and that acuity is mostly limited to good instances of the exemplars in question (an agent might not be able to identify far-from-typical A's correctly), and favourable sorting conditions. This yields a problem that is very similar to one discussed earlier, namely the problem of concept combination as it arises for Prototype Theory: pet fish are neither typical (i.e. good) instances of the category 'pets', nor of the category 'fish'. I trust the explanation given in section 6.11.2, when this problem arose before, suffices to support the claim that SToCC is not harmed in any significant way by the 'compositionality'-argument.

Fodor's third and final attack on BCP, the 'Circularity'-argument, says that both the 'sorting' and 'inferring'-conditions of concept possession yield vicious circularities. On BCP, being able to sort all triangles from non-triangles means having the concept 'triangle'. But what are these sorting-actions based on? Surely, on the conceptual knowledge that one has, which would need to include the concept 'triangle', or a conceptual equivalent (such as 'closed trilateral'). Circularity also arises, argues Fodor, for 'inference' as used by BCP. An inferential grasp of 'conjunction' necessarily presupposes the understanding of the 'and'-operator, it seems - any and all

definitions that attempt to infer it without already using it (implicitly or not-so-implicitly) fail.

I suspect much of what Fodor finds problematic in both these cases derives from his insistence that concepts are bound up with - or constituted by - representations, defined *internal* to the mind. So runs the argument: possessing the concept 'conjunction' is exemplified in 'conjunction-understanding behaviour', which includes inferences to recognising the content of the conjunction-concept, but that already presupposes the possession of that concept as an internal representation, with a particular function in mental processing involved in exhibiting 'conjunction-understanding behaviour'. And there we have circularity, says Fodor.

I can concede Fodor's circularity-argument as applied to BCP without problems, for according to SToCC, concepts are *not* defined solely in internal terms - rather, conceptual possession is defined in embodied and embedded terms. For instance, sorting on the basis of a particular concept is actually sorting on the basis of bits of knowledge, observed and remembered behavioural profiles, and environmental cues that form the extension of the concept's enslaver, and the inferences based on that enslaver's contents. In other words, exhibiting a specific kind of behaviour as a criterion for concept-possession (e.g. carrying out a sorting task successfully) is not circularly dependent on having said concept, but rooted in the *narrative* that forms the embodied and embedded substrate of the *accountability* involved in having a concept (see section 6.6). This narrative can exhibit a marked difference between the case in which an agent possesses (and acts upon) the concept 'triangle' and the case in which the concept 'closed trilateral' is involved, for instance in terms of the acquisition history for either concept. That is, learning about triangles during a math class need not imply the acquisition of an *explicit* definition of such objects in terms of the concept 'closed trilateral'.

6.11.5 - SToCC vs. Conceptual Role Semantics

There are a few similarities between SToCC and Conceptual Role Semantics (CRS). CRS arises from a functionalist approach to the mind: a weak form of CRS aligns, to a certain extent, with analytic functionalism:

... (a representational state) "is *meaningful* (i.e. has *some meaning or other*) by virtue of the fact that it plays a certain role in a person's psychology." (Block 1998)

So, according to CRS, the meaning of a concept is determined by its psychological role, hence (in contrast with more behaviouristically inclined theories) talk of internal states is explicitly included in explanatory accounts. A stronger form of CRS expands the characterisation of weak CRS above by positing that not merely any content whatsoever would be characterised by its role, but that there is a connection between specific mental content and the role it might play in certain processes.

In SToCC, a concept's content is also defined by its use (structuring by the associated account's narrative situatedness, which in the case of more carefully defined theoretical concepts can consist of a particular theory's definitions corollaries), but by virtue of conceptual space's groundedness in perceptual performance, there is a basis of *intrinsic*, neurophysiologically informed content: the Neurophysiological Yield (NPhY; see section 5.2 and note 31).

Block (1998) highlights a common objection to CRS:

"CRS is often criticized from the point of view of truth-conditional theories of meaning. If the meaning of a sentence is its truth conditions, then the meaning cannot be its conceptual role."

In SToCC, the conditional of this latter statement is called into question, by claiming it appears profligate to speak of *truth*-conditions[NOTE 65]. Rather, SToCC would support speaking of *possession-conditions*, but not in a binary fashion (i.e. either you have a concept or you do not, without intermediate options); even better would be *appropriateness-of-use conditions*. Appropriateness-of-use judgments define a domain of correct usage, and a subject satisfies the possession-conditions for a specific concept if the way he uses said concept (either in speech, thought or act) falls within this domain, i.e. if the sentences containing the concept the subject utters are understood to be correct (or at least appropriately so) by the other users of the concept, if they accept the justificatory account. This creates a mutual-attunement-dynamic between these concept-users: for shared theories, such as scientific ones, this is a communal, hence socially embedded domain. Obviously, the ultimate touchstone is the real world: the range of ways in which a concept is used by a group of people should be defendable with proper empirical arguments.

In short, SToCC does not, in the end, endorse any one approach in the philosophy of science regarding the truth of theories. Instead, the weaker notion 'appropriateness-of-use-judgments' is to be understood as a mild suspension of judgment on this matter. However, the underlying thrust of the account is that the modern scientific method of empirical, argument-based confrontation of differing theories is the correct kind of activity to improve our knowledge of the world. Whether any one opinion is truth (ultimate and eternal) and what the criteria for this would be are not questions SToCC aims to confront. Quite the opposite: SToCC is intended to forge a semblance of unity in a chronically, possibly even *fundamentally* fragmented field of research. After all, the notion 'colour' can be understood in many different ways, ways that, in terms of actual content, are not in an obvious sense connected, except that they appear to refer to various components of a single, highly complex process. SToCC is intended to help dissolve this tension between a fragmented epistemology and an apparently unified ontology.

Because in SToCC the meaning of a concept depends, to an important extent, on the associated, justificatory account, the need to find analytical bedrock for a concept's meaning can be relinquished. In everyday situations, agreeing to disagree beyond a particular granularity suffices, but even when more robust justification is required, for instance in terms of an actual scientific theory, this non-reductivist inclination can be upheld. This is due, to an important extent, because of the properties of the theories that might be called upon for support. Any scientific theory *is* explicitly embedded in a web of propositions, and, if one traces the connections far enough, to the totality of human knowledge as well as its axioms, but it is artificially walled off from non-relevant propositions by virtue of its definitions (and their corollaries). Theories (almost?) always concern artificial segmentations - often even *idealised, abstract and/or ceterus paribus models* - of the interconnected totality of reality. Obviously this segmentation is usually modeled after intuitively natural categorizations, but to suppose that science always or even usually cuts nature exactly at its joints is disputable. This discrepancy of a theory with the actual situation is already accomodated for in the appropriateness-of-use-judgments: in the vast majority of practical cases, there is no need for an boundlessly exact fit of the actual encountered situation with the textbook example of a particular phenomenon.

In that sense, SToCC's solution is similar to one of Block's suggestions:

"A third approach to accommodating holism with a psychologically viable account of meaning is to substitute close enough similarity of meaning for strict identity of meaning. That may be all we need for making sense of psychological generalizations, interpersonal comparisons, and the processes of reasoning and changing one's mind." (Block 1998)

Gilbert Harman defends a specific kind of CRS: Nonsolipsistic Conceptual Role Semantics. He says:

"(Nonsolipsistic) conceptual role semantics may be seen as a version of the theory that meaning is use, where the basic use of symbols is taken to be in calculation, not in communication, and where concepts are treated as symbols in a 'language of thought'. Clearly, the relevant use of such 'symbols', the use of which determines their content, is their use in thought and calculation rather than in communication." (Harman 1998)

In (stark) contrast, SToCC steers clear of a 'language of thought'-hypothesis of the computationalist kind. At some level of description, SToCC can be about things (i.e. concepts and associated notions) that might, in some sense, be represented in consciousness, and might exert influence in a subconscious way in some cases. Whatever explication is to be awarded to this string of what some might call 'weaselers', SToCC is quite clearly *not* about symbols or somesuch entities that are supposed to reside at the most basic and fundamental level of thought. Rather, SToCC is about embodied

agents acting in an environment, and concepts are a way of describing certain aspects of that interaction dynamic[NOTE 66].

Harman continues:

"Concepts and other aspects of mental representation have content but not (normally) meaning (unless they are also expression in a language used in communication). We would not normally say that your concept of redness meant anything in the way that the word 'red' in English means something. Nor would we say that you meant anything by that concept on a particular occasion of its exercise." (Harman 1998)

SToCC holds that concepts actually *mean* something, in the sense that they involve the behaviour-directed attitutude of an agent towards some object, process or state of affairs in a specific way (namely, the way locked in by the associated narrative, as contextualised in the utterance or think-act of the proposition the concept is a component of). An important part of SToCC is the idea that a concept is often not, ultimately, a unitary notion, even though it is often used as such. A 'concept of redness' means whatever the associated justificatory account says it is, in the relevant context. 'Red' can mean something only by virtue of its embeddedness in theories and the world as well as the fact it is used by embedded 'animals' in a particular way. 'Red'-the-word means something by virtue of the meaning of the associated concept, which means something itself by virtue of its associated justificatory account[NOTE 67]. Content, in SToCC, is never solely intrinsic, and perhaps this is exactly what Harman's one-place CRS, as opposed to Block's two-place CRS, is about, for Harman might have to agree that content depends on something external:

"The moral is that (nonsolipsistic) conceptual role semantics does not involve a 'solipsistic' theory of the content of thoughts. There is no suggestion that content depends only on functional relations among thoughts and concepts, such as the role a particular concept plays in inference. Of primary importance are functional relations to the external world in connection with perception, on the one hand, and action, on the other." (Harman 1998)

This sounds quite compatible with SToCC: a concept is an embodied/embedded agent's disposition towards action; it is about some object, state of affairs or process, and this intended entity is needed to specify what the concept is.

In SToCC, there might be some aspect of conceptual content that is intrinsic: the neurophysiological basis of perceptual judgments (the Neurophysiological Yield). But that kind of content on its own does not a concept make (let alone a structured web of concepts). This neurophysiologically defined structure of saliencies needs to be embedded within a context of use for it to be in any way relevant, because only then can it serve to underlie meaning.

We hit upon an important difference between Harman's account and SToCC, as Harman says:

"(Nonsolipsistic) conceptual role semantics asserts that an account of the content of thoughts is more basic than an account of communicated meaning and the significance of speech acts. In this view, the content of linguistic expressions derives from the contents of thoughts they can be used to express." (Harman 1998)

Of course the concepts and their contents underlie the speech acts, but the kind of relation of the speech act to the intended object (or process, or state-of-affairs...) is not radically different from the kind of relation of the concepts to that same object, in part because, just like speech acts, concepts are contextual, inherit their meaning in use, refer to something, and so on. This is why differentiating acts (speech, thought, bodily) and concepts on the basis of the way they are connected to that outward object (or process, or state-of-affairs...) will not work.

A suggestion for a different strategy, allowed by SToCC, could be to assert that appropriateness-of-use judgments allow different uses of a concept to actually be tokens of the same thing, whereas acts (speech, thought, bodily...) are more obviously anchoured to a specific time and place: they are 'happenings' rather than recurrent uses of the same (or a similar) schema (i.e. concept).

In summary, SToCC differs from Block's CRS because of its ability to parry the 'truth-conditions'-counterargument, and from Harman's Non-solipsistic CRS by being, in general, much less cognitivistically inclined.

6.12 - Intermediate Conclusion

Based on the results of this intermediate evaluation, I claim that SToCC offers an appropriate account of concepts: SToCC's notions enslaver and granularity *can* provide an account in which concepts are cognitively economical, yet informative qua inferred content. This inferential structure is expressed in the structure of conceptual space, and part of this structure is explained by the mechanism of conceptual space splitting: a splitting sequence provides a historical explanation of why certain complex concepts are compactly represented at a particular granularity level.

Furthermore, these mechanisms and properties are linked to the agent's embodied and embedded nature; two components exemplify this link most explicitly: (1) an enslaver is derived from perceptions and assessments which occur in the embodied and embedded perception/interaction of an agent; (2) the precise trajectory of conceptual space splitting depends on the way in which the agent functions in his environment, i.e. the kinds of things he perceives, learns, does, combined with how his environment reacts to him.

This means that, clearly, SToCC *needs* a broader framework of embodied/embedded cognition to explain why a particular concept is the way it is, e.g. why a particular concept has this or that enslavement hierarchy, or behaves a certain way under the context-driven exploration of a granularity gradient, without the agent explicitly holding a concept's associated theory. More specifically, I have to say more about how conceptual space is linked to the environment, to the agent's behaviour and to the biomechanical properties of his body. The model to be developed next, called the *Radicality Manifold*, provides a way of understanding the interactive, embodied/embedded dynamics which SToCC needs for support.

= =

[SUMMARY of chapter 6 AND PREVIEW]

In this chapter, the idea of 'colour' as a complex concept (which has applications in different contexts which cannot be reduced, without remainder, to a single definition: the 'combined' colour concept incorporates mutually exclusive notions, hence was characterised as a 'superposition') gave rise to the idea that a concept in general is context-dependent. This lead to an $E_{(i)}C$-appropriate concept definition: a concept is a structured behavioural disposition of an embodied and embedded agent - behaviour, in this case, includes cognition and locution. Conceptual space, then, is an expression of the structure that is inherent to the inferential connections between the kinds of behaviour an agent might exhibit as *justification* of his use of a particular concept. Concepts are interrelated along inferential lines, and are informed by narratives, which are built out of the agent's own experiences, memories and ideas.

Conceptual enslavement is the phenomenon that such a narrative as it informs a concept can have a particular centre of gravity - experiences and such which form a comparatively great contribution to the meaning of the agent's concept, and which will be offered more readily than other ideas in explanation of his use of a particular concept. The *granularity* at which a particular concept is explained may vary from situation to situation, and will help two agents come to the realization that they share a concept, even though one of the two might have a much deeper understanding of said concept, if he were pressed to explore those depths. In many everyday situations, however, such detailed accounts are not necessary, and a low-grain recognition of the other's contextually appropriate concept-use is deemed satisfactory. This practice highlights an important aspect of concepts: a concept is what you do/say/think in a particular context, to an important extent as appraised by a conspecific. Having concepts is, in part, getting the acknowledgement from others that you did, in fact, use said concept in an acceptable fashion (see section 9.2 for more on this). These ideas together yield *Superposition Theory of Complex Concepts* (SToCC),

where superposed complex concepts were understood as special cases of a general concept theory.

Given the idea of conceptual space (expressing all concepts a specific agent has) as a spectrum ranging from basic sensorimotor acuity to complex, abstract ideas, it is possible to sketch the development of a conceptual system. In this chapter, I outlined a four-stage process, incorporating the following stages:
[*Stage 1*]: sensorimotor apprehension of motion;
[*Stage 2*]: correlation of sensorimotor knowledge and linguistic encoding;
[*Stage 3*]: embodied and embedded crossmodal mapping;
[*Stage 4*]: correlation of embodiment and abstraction

A description of the continued development of conceptual space, once it is actually in place, along the lines set out in chapter 5 (in which perceptual space segmentation was described), was also provided. More finely-grained and sometimes even new concepts can emerge via conceptual space splitting, which involves the progressive segmentation of conceptual space as an *embedded manifold*. Understanding conceptual space as an embedded manifold is to say that conceptual space can be conceived as being composed of interlocking regions expressing various subconcepts and inferential connections, ever more detailed at ever finer granularities. Some of the most basic divisions of conceptual space depend on the categorizations enforced by the properties of our body and our senses.

An intermediate evaluation of SToCC - comparing it to Prototype theory, Theory theory, Jerry Fodor's view and Conceptual Role Semantics - suggested that it compares favourably to those existing concept theories. An important comment to repeat here concerns the difference between prototypes and enslavers: prototypes collectively lock in the *definition* of a concept, whereas an enslaver enables *inference* towards ideas and behaviour that are appropriate for someone professing to have a particular concept. However, in line with the discussion in the chapters before this one, it is still apparent that a more detailed description of the interrelatedness of conceptual space ('C-space') with bodily properties, sociocultural environmental properties and physical environmental properties (respectively: M-, S- and P-space) is needed: why is conceptual space structured the way it is? How can we say more about the ways in which an embodied agent is embedded in his environment, and how this is reflected in his concepts? Chapter 7 is the first step towards answering these questions, providing an $E_{(\,i\,)}C$-appropriate account of representation to help define that relation between body, social environment, physical environment and concepts.

= =

[7 - The Radicality Manifold: Preliminaries]

7.1 - Introduction

In the previous chapters, I have attempted to synthesize $E_{(i)}C$-perspectives from already raging discussions about the ecological and socio-linguistic agent-environment-interactions involved in colour perception. Because the philosophy of colour perception constitutes a microcosmos of the philosophy of mind, touching upon many of the major themes of that broader domain but in a more compact manner, this preparatory work allowed me to construct a more general, $E_{(i)}C$-compatible account of concepts: Superposition Theory of Complex Concepts. It is my intent to use this theory of concepts, in conjunction with the behavioural planning field that was discussed in section 3.2, to yield a few suggestions about how to think about concepts and higher cognitive abilties in an $E_{(i)}C$-perspective.

These topics would require another book (at the very least) to address properly; therefore, in the pages to come I can only offer suggestions, sketches and hypotheses. What I do hope to accomplish, is to demonstrate that even though a lot of research is still needed, the model developed below, the *Radicality Manifold* (RM), hopefully constitute a modest nudge towards a more substantial understanding of $E_{(i)}C$, as well as concepts.

I will start with developing a characterisation of low-level content, i.e. the kind of content present at the lower reaches of the conceptual space spectrum. To do this, I will use the work of Dan Hutto as a springboard. This might come across as odd; after all, for Hutto, *low-level content* is an inherently contradictory notion, as it implies a kind of reification of mental entities that he feels lies at the root of many current problems in the philosophy of mind: whatever is 'low-level' cannot be 'content', at least not content 'inside the head'.

Still, I will offer some suggestions for conceptions of *content* and of *representation* which can be made compatible with the (quite substantial) subsection of Hutto's account that I wish to salvage. In line with the critique Prinz and Barsalou (2000) level against the defenders of dynamical systems theory as applied to cognition, the tenor of the sections to come will be that supporters of $E_{(i)}C$ (and even $E_{(A)}C$) are under no strict obligation to do away with any and all talk of content and representation. On the contrary, if sufficient care is taken to avoid the pitfalls of the old cognitivist approaches, these conceptions can *strengthen* rather than weaken the $E_{(i)}C$ programme.

7.2 - Radical Enactivism

It is to Dan Hutto's philosophy that turn to once more (I visited his Narrative Practice Hypothesis earlier, in section 6.6), because his opposition to talk of *content* being involved in cognition is illuminating. Hutto (e.g. 2006) calls his position *Radical Enactivism*, and as such he allies himself with other

philosophers and scientists working in the enactivist tradition, but with a critical inclination: many enactivists are, apparently, not *radical* enough in following through on their own principles. That is, his theory embodies an attempt to purge the last representationalist remnants from the enactivist program (recall that Thompson's theory of colour is enactivist). One of the formative papers of the enactivist movement, O'Regan and Noë (2001), contains several claims and explanations which Hutto invokes to support his case that there is still some cognitivist ballast left to exorcise. Consider the following quotations, which come dangerously close to containing references to kinds of *knowledge* that are supposed to *mediate* between sensory input and behavioural output:

"In what does your focussing on the red hue of the wall consist? It consists in the (implicit) knowledge associated with seeing redness: the knowledge that if you were to move your eyes, there would be changes in the incoming information that are typical of sampling with the eye; typical of the nonhomogeneous way the retina samples color; knowledge that if you were to move your eyes around, there might be changes in the incoming information typical of what happens when the illumination is uneven, and so on." (O'Regan and Noë, 2001)

"...seeing is a skilful activity whereby one explores the world, drawing on one's mastery of
the relevant laws of sensorimotor contingency." (O'Regan and Noë, 2001)

Implicit though the 'knowledge' and 'mastery of laws' might be according to O'Regan and Noë, Hutto, an avowed Wittgensteinian, claims this is not nearly radical enough. As stated, he baptised his own theory *Radical* Enactivism, to stress the fact it removes any and all hints of representationalism at the sensorimotor level.

Hutto defends a distinction between *basic visceral responding* and *linguistically mediated thought*. The level of basic visceral responding is the level at which most of the sensorimotor activity takes place that is central to enactivist accounts, and it is the confused description of some enactivist writings, still involving analyses in terms of mediating knowledge and representation-like entities, that Hutto targets. At this basic level, there can be no talk of content. There are no internal states (e.g. symbolic representations) at work here - rather, the dynamics of these processes should be understood in terms of *contentless intentional directedness*.

The bedrock on which this view is founded, is biosemiotics[NOTE 68], in particular Ruth Millikan's ideas involving *intentional icons*, i.e. the kinds of signs and 'representations' (Millikan's word, not Hutto's, but the gist of the latter's claim is the same) that play a role in an interlocking action/interaction dynamic of agent and world.

These are the features of intentional icons:

"1. They are relationally adapted to some feature, object or state of affairs.
2. The relation described in (a) can be characterised by means of a mapping rule.
3. They have the direct proper function of guiding co-operating (consumer) device(s) in the performance of its (or their) direct proper function(s)." (Hutto, 2006)

Millikan describes these 'representations' as *pushmi-pullyu's*: signs that are both descriptive (stating what is the case) and directive (stating what is supposed to be the case, e.g. what should be done). The proper function of a hen's call to her chicks, for instance, is to direct the chicks towards food; the descriptive content would be something like 'there is food right here!', the directive content would be 'come here and eat!' (Millikan, 1996).

Another example concerns bee dances, used by scouts to indicate sources of food to conspecifics that were left behind in the hive. The point that Hutto wants his readers to take home is that these dances are not representations of the location of nectar, nor do they *contain* any *information* whatsoever. Rather, following Millikan he claims the dancing and spectating bees enact a specific interaction dynamic that has evolved for a particular purpose: the dance has the proper function of evoking a particular kind of response, namely a sufficient number of bees leaving the hive, flying to the food source and bringing the nectar back home. As Millikan says:

"Bee dances, though (as I will argue) these are intentional items, do not contain denotative elements, because interpreter bees (presumably) do not identify the referents of these devices but merely react to them appropriately." (Millikan, 1984)

Hutto's (2006) claim, following Millikan, is that these signs are patterns of interaction dynamics that evolved to evoke a response, and nowhere in or during this process does there need to be an explicit representing, mastering, processing or decoding of rules, laws or any other kind of *content* that would somehow be present in the bees' dances. The almost-rhetorical question then becomes: why would our basic sensorimotor activity be different? Human intentional directedness, as well as the causes of our basic sensorimotor responses are to be understood in terms of biologically proper functions - in terms of these causally interlocking processes amongst conspecifics that does not require the encoding and decoding of symbolic representations to work.

As announced, the second issue to be discussed involves *concepts and content*, especially the idea (defended by Hutto) that content has no place in descriptions of basic sensorimotor interaction. If coherent, this idea could prove detrimental to SToCC/RM, which includes the idea of a conceptual spectrum that is supposed to include basic sensorimotor processes as well as complex cognitive activity.

To get a tighter grip on this claim, regard Tim Crane's view, which Hutto criticises in (Hutto, 2006). Crane characterises the intentionality of a sensation as a relation holding between a subject, an intentional mode and a specific intentional content. For instance:

"the content of the sensation is *that* one's ankle hurts, the object of the sensation is the ankle (apprehended *as* one's ankle) and the mode is the hurting." (Crane 2003)

This content need not be propositional. Crane says it is 'what one would put into words, if one were to have the words into which to put it' (Crane 2003).

There is something to be said in support of Hutto's worry that this opens the door to an intrusion of the conceptual (in terms of Hutto's apprehension of concepts, i.e. linguistically specifiable) on the nonconceptual realm. Hutto says that talk of content is fine when characterising conceptual aspects of experiential modes, but not when the nonconceptual aspects are at issue; these aspects are are defined as not involving judgments of any kind. Hence, the basic capacities of experience are non-conceptual and *not* content-involving

It should be clear that Hutto utilises a specific notion of 'conceptual', one that automatically implies content. What is more, Hutto endorses a strongly Fregean conception of 'concept', where a concept's content is (or should be) linguistically specifiable. Anything that is not so specifiable, cannot be conceptual; for Hutto, the notion 'nonconceptual content' is inherently contradictory.

The notion I wish to defend concerning concepts and content is not Fregean, but Hutto's insistence on these issues *does* force me to be much more specific about what it is that I am trying to say. An important aspect of my message is that the phrase 'having a concept' in the provisional characterisation of concepts provided in section 6.3 is actually a rather *misleading* way of speaking. In SToCC/RM, the phrase in question denotes 'having a capacity, in a particular context'; more in particular, the capacity to provide an inferred account and/or behave in an appropriate fashion, when pressed to do so. These concepts-as-capacities are also intended to be present at very basic levels - levels where Hutto claims there can be no talk of content whatsoever. Is it possible to reconcile these views? And if this cannot be done, what position is inherent to SToCC/RM concerning the content of concepts, even basic ones?

The first step towards an answer is as follows: it is very important to note that the 'having' of a capacity is quite different from the 'having' of a car, or of blue eyes. Nonetheless, I would like to claim that we at least (I am not sure about other animals) are capable of assuming a perspective on our capacities, of reflecting on their structure and properties. That is, we can

have a relation to our own concepts-as-capacities; does this lead to the kind of reification and content-ascription Hutto hacks away at so adamantly?

I want to say: yes and no. Yes, I suggest that there should be (at least some) talk of content, but no, this need not lead to unwarranted reification of concepts.

In SToCC, the relation one can have to one's own capacity to do/say/think X (the mark of concept-'possession' on this theory) is mediated by the inferred (narrative) account, and its behavioural consequences. I suspect (but will not argue this suspicion extensively at this time) that this ability is at least socially, and possibly also linguistically mediated: the ability to engage in social interaction and utilise language might be a precondition for generating such inferred accounts, even the non-linguistic ones. As such, the 'having' of concepts involves content, but I believe this claim is not in conflict with Hutto's ideas, for these narrative accounts can be placed squarely in the linguistic/narrative realm, the relevance of which he himself also advocates.

However, it is at the more basic end of the sensorimotor spectrum that the conflict between Hutto and SToCC is supposed to emerge. My claim is that the capacities at play here also involve content, but in a very particular way. To support this claim, I need to cover quite a bit of ground: the topics to be discussed next include conceptual content (obviously), but also representations, truth conditions (again!), and the realization of properties that the RM-account is capable of providing.

7.3 - Representation

First, for a somewhat clearer notion of what 'representation' actually is, we can look at Menary (2006), who presents the following general definition of 'representation':

"1. A token vehicle Φ is a representational vehicle when it has properties that can *potentially* be exploited by a representational consumer. For example: Φ is salient because it is reliably correlated with an object/environmental property X, or with objects/environmental properties X, Y, Z...

2. Φ has a representational function when its salient features are exploited by some consumer Ψ. For example: Φ has the function of representing X for consumer Ψ, because Φ is reliably correlated with an object/environmental property X.

3. Φ represents X for consumer Ψ in the performance of some biological function." (Menary 2006)

This is a kind of representation that Bennett and Hacker (2003)(see note 1) described as representation which involves the correlation of two states

(e.g. one internal, the other external); these states are mutually attuned due to being causally linked in some way.

Peschl and Riegler (1999) discern several main structural variations on this theme. Defining the two poles involved in the correlation as *Realität* (R; the real world 'as it is in itself') and *Wirklichkeit* (W; the constructed world 'in our heads'), these are the options they suggest:

* Naive imaging: *W = R*
The 'real world' and the internal representation display a one-on-one mapping.

* Classical representational theory: *W = f(R)*
Internal representation W refers to external reality R, but is distorted in some fashion via function *f(x)*.

* context-dependent representation: *W = f(R, O, C)*
internal representation W emerges in interaction with reality R, but is modulated by properties of the observer O and cultural influences C

* Self-referential Representation: *W = f(W, E, P)*
Internal representation W depends, via function f(x), on the structure and content of W itself, background experiences E, and perturbation events P. The absence of R indicates the operational closure of the cognitive system: the representational dynamic self-organises, and external input (sensory stimuli) are merely perturbations of that dynamic.

At first glance, the causation/correlation process that lies at the basis of Menary's definition, understood in a context-dependent way as outlined by Peschl and Riegler, shares some properties with to the biosemiotic approach defended by Hutto (see section 7.2 above). However, it is the interpretation or operationalisation of such biosemiotic correlation in terms of representations as mental 'objects', the features of which need to be 'exploited', and this exploitation process involving or yielding 'information' (as mental content), that affronts Hutto.

7.4 - Representation and $E_{(A)}C$ [NOTE 69]

As has been established, many forms of $E_{(i)}C$, including the $E_{(A)}C$ of Thompson (see chapter 4) and the dynamicism of Thelen and colleagues (see chapter 3), are not in the least bit fond of explanations of cognition that involve reference to representations. The invocation of representations by Fodor (1975), and in particular by Marr (1982) in his account of visual perception, is diametrically opposed to the $E_{(A)}C$-approach, it is often claimed. Is this really true?

Not all $E_{(i)}C$-supporters are morbidly opposed to incorporating the notion 'representation' into theories about cognition. Clark (1997) offers a version of representationalism that holds the middle ground between, on one

extreme, the kind of explicit symbol-based representation that, for instance, Gibson (1979) argues against (but which, I believe, isn't a very popular position any longer anyway, if it ever was), and at the other end of the spectrum the kind of non-representationalism he and like-minded dynamicists (such as Thelen et al.) argue in favour of. Clark says (1997, p. 147):

"(. . .) let us call a processing story representationalist if it depicts whole systems of identifiable inner states (local or distributed) or processes (temporal sequences of such states) as having the function of bearing specific types of information about external or bodily states of affairs."

This 'definition' offers room for a dynamicist connectionism to still use internal representation. The possibility to use representation in explanations is due to what Clark calls 'representation-hungry' problems: situations that might arise in the life of an animal, which require some modicum of representation to be dealt with successfully – in these cases, an explanation using representation is the best one available. These involve situations in which Haugeland's main criterion for the existence of representation is satisfied:

"(a system uses internal representation if) it (coordinates) its behaviors with environmental features that are not always ''reliably present to the system''." (Haugeland, 1991, p. 62)

Clark mentions two possible scenarios that satisfy this condition: (1) reasoning about states of affairs that are absent or counterfactual, or do not exist (thinking about the past or the future); and (2) selective sensitivity to complex, multi-interpretable stimuli.

Regarding Clark's first category, it remains to be determined what kinds of 'reasoning' belong to this particular category, and where exactly one must seek the boundary between cases demanding explanations involving representation, and cases that can do without it. Van Rooij, Bongers and Haselager (2002) build an interesting case for a non-representational explanation (using dynamicist models) of a type of imagined action (determining whether a rod, handed to the test subject, would be of sufficient length to reach an object, without actually performing the task) that could be claimed to lie within the most primitive reaches of the domain Clark identifies. If the model by Van Rooij et al. continues to hold up under strict scrutiny, this would require a sharper definition of Clark's concept 'representation-hungry', and would imply that the subset of situations to which that term could be applied is smaller than Clark assumes.

With the research of Van Rooij et al. to weaken the force of Clark's case for the prominence of the first category, I would suppose the suite of situations belonging to Clark's second category presents a somewhat stronger case for the necessity to utilise representation in explanations. Furthermore, the associated representing ability would involve the capacity to construct and

use abstractions, and would therefore touch upon much of what is characteristic of human cognition. Clark offers the examples of selecting all the valuable items in a room, or all items belonging to the pope. In such cases, the selection criterion resides on a rather abstract level that has little, if anything to do with the directly observable physical characteristics of the objects. This means the selection process would necessitate internal representation: keeping in mind (rather literally) which feature is relevant and how to select items based on it. Clark tones down this conclusion by stating such a representation is not necessarily an (easily?) identifiable activation pattern in a specific brain region in the sense that old-fashioned computationalists might want to endorse. The representation in question could very well be highly complex, both in temporal structure and in physical instantiation and distribution. Furthermore, representations often, perhaps always serve a behaviour-guiding role, and from this point the step towards an account favourable to some of the dynamicist's central tenets is not so big any more.

After all, subsequent development of these ideas by Clark takes place in the context of his overarching project, in which he explicitly endorses an embodied account of cognition, signifying a slight additional shift towards the dynamicist position. Clark explicitly notes the compatibility of the attack on certain aspects and interpretations of computationalism by Thelen (1995) and Thelen and Smith (1994) with more sophisticated versions of theories from the computationalist programme.

Thelen's (1995) rejection of computationalism is not, out of necessity, incompatible with the kind of view Clark endorses. Thelen attacks two theses:

(1) Piaget's claim involving innate knowledge (Clark, 1997, p. 155: "development is driven by a fully detailed advance plan");
(2) the textbook (and rather coarse-grained) characterisation of computationalism (once more Clark 1997, p. 155: 'adult cognition involves internal logical operations on propositional data structures')

As alternative suggestions, Thelen and associates put forth the following two theses (adapted by (Clark, 1997, p. 155):

(a) "Development (and action) exhibit order which is merely executory. Solutions are ''soft assembled'' out of multiple heterogeneous components including bodily mechanics, neural states and processes, and environmental conditions (Thelen and Smith 1994, p. 311)"
(b) "Even where adult cognition looks highly logical and propositional, it is actually relying on resources (such as metaphors of force, action, and motion) developed in real-time activity and based on bodily experience. (Thelen and Smith, 1994, p. 323; Thelen, 1995)"

In a sensibly moderate interpretation of these theses, there is nothing here to contradict an embodied, dynamicist connectionism with a properly dosed use of representationalist explanations.

Clark defends his program from possible criticism deriving from the work of Gibson (1979), and this defense might also serve to deflect possible counterarguments of dynamicists, and serve to indicate wherein the moderation mentioned needs to consist. For Gibsonians and Van Gelder's dynamicists alike, internal representation will only cause problems as an explanatory component if it is understood in a particular way, namely as 'rich, action-neutral encodings of external states of affairs' (Clark, 1997, p. 172). The emphasis, present in the original text, isolates the crucial (dynamic!) element. A more general concept of internal representation as "inner states, structures, or processes whose adaptive role is to carry specific types of information for use by other neural and action-guiding systems" (ibid.) is much easier to reconcile with the embodied approach of, for instance, Thelen and colleagues, partly because of its relative ontological neutrality, but mainly due to the emphasis on its role in the guidance of (embodied) action. This general concept of internal representation does appear to run the risk of being used as an explanatory panacea by those inclined to support the notion. Partly due to its generality, philosophers might feel less inhibited in using this kind of representation as silly putty to close gaps in their theories. Dosed usage of this explanatory possibility is recommended, and I get the impression both Clark and Thompson and colleagues (with their neurophenomenological method, to be discussed below) are careful enough.

But to repeat the claim above, there is nothing in Thelen et al.'s rejection of what they perceive computationalists to claim to contradict an embodied, dynamicist connectionism with a properly dosed use of representationalist explanations. That is, I would say there are certainly possibilities for combining dynamicist and representationalist approaches in a fertile manner. Van Rooij et al. (2002) note that the main difference between an explanation involving representation on one hand and an explanation using the DST toolbox on the other, is assumed to be that the former offers a mechanism, whereas the latter can merely describe. Their solution to this dichotomy is to claim that both representationalist and dynamicist explanations provide mechanisms, in the sense of specifying the constraints the underlying physical process is subject to.

I would maintain that neither account is complete, the representationalist perspective offering a top–down, somewhat metaphorical view of the phenomenon, and the dynamicist approach a more abstract, behaviourist account, while neither actually touches upon the phenomenon and its mystery itself, leaving a gap between the two approaches. The big question would be: how can the brain generate cognition? How does that work? Despite this shortcoming, the research conducted by Van Rooij et al. does enforce the highly important lesson we encountered earlier: when one

wishes to use representation to explain some cognitive phenomenon, make sure this is absolutely necessary.

7.5 - Types of Representation

Clark's discussion appears to centre on the relevance of the more-or-less classical idea of representation as an internal symbol-like entity (such as a memory) standing in for something external, for instance when that external object is no longer present for immediate perception.

My suggestion will be that representation can be highly useful in explaining cognition in a broader sense, *if and only if* representation is understood in the appropriate fashion. Prinz and Barsalou (2000) appear to agree. They note that traditionally, representations are conceived as context-invariant, disembodied and static; it is exactly this orthodox conception of representation that is unacceptable to many $E_{(i)}C$-supporters, defenders of $E_{(A)}C$ in particular. However, at least regarding the dynamicist form of $E_{(A)}C$, Prinz and Barsalou suggest that 'dynamic systems theory and perceptual symbols theory are complementary. They can work in concert to describe different aspects of cognitive systems.' (Prinz and Barsalou 2000). A very similar conclusion was reached in section 4.5, where Thompson's enactivism and Shepard's ecological computationalism pertaining to colour perception were claimed to complement eachother in a rather interesting fashion.

Prinz and Barsalou invoke the account of representation developed by Fred Dretske (see below) to bolster their own claims. The general idea of that account is to parse representation in teleological/informational terms: a state represents some entity if the state and entity exhibit some form of covariance, and the state actually has the function of somehow containing information about that entity. Prinz and Barsalou state that if representation is explained along Dretske's lines, this informative covariance-relation, is capable of exhibiting the contextual sensitivity, embodiment and time-dependent dynamics $E_{(i)}C$-supporters, including dynamicists submit are essential aspects of cognitive systems. Please be advised that the notion 'information' is not a neutral term - see section 7.6 below for more on this topic. Here, I suggest we understand 'information' not in contentful terms, but as an external-to-internal causation of particular structures or forms on an agent's internal dynamics.

But first, I suggest we take a look at the theory developed by Fred Dretske, which is one of the most important views that is habitually associated with the 'correlation view' of representations, to determine whether this internalistic, cognitivistic approach is the only possible one, or whether there are more constructive, $E_{(i)}C$ and $E_{(A)}C$-friendly options available.

Dretske's (1988) influential theory about representation contains distinctions between different kinds of such representation relationships, which he has

labeled Types I, II and III; only Type I representation is explicitly symbolic in the sense disqualified by most $E_{(i)}C$-supporters.

(Conventional) Type I Representation: The representational powers of the elements of Type I systems are not intrinsic, but derive from the creators and users of said systems, e.g. coins and bottles might represent a sequence of events during a sporting match by way of their relative positions and movements, as manipulated by an agent who claims to remember this sequence, and wishes to enact it for his audience. The objects are the agent's representational instruments. Dretske calls these representational elements *symbols*. These systems are doubly conventional: their function is assigned to them by us, and the actual functioning is due to our manipulations and actions.

Given this description, it would indeed be silly to suppose that something like this kind of representation would occur 'in the head' or 'in the mind': this mode of representation appears to require a shared attentional focus, a particular imposition of structure, and an interpretational act (someone manipulating the objects-as-symbols, placing them in particular relations, and another person regarding these symbols and attempting to recover the referential 'content'). These are *external* and *social* rather than *internal* and *mental* occurrences.

(Conventional) Type II Representation: In Type II systems, *natural signs* perform the representational function that symbols perform in Type I systems. Natural signs are, for instance, animal tracks and approaching thunder clouds, and they signify what they do (the presence of animals at this location in the recent past, or the increased likelihood of rain falling here in the near future) quite independently of our interpretatory acuity and action. Symbols (which we find in Type I systems) mean whatever we say they mean, whereas signs, which figure in Type II systems, possess natural (i.e. non-conventional) meaning we might (or might not) pick up on. The connection between the sign and what it signifies is one of the sign *indicating* the presence or occurrence of the signified, consisting of a correlation that is lawful and persistent. In Type II systems, these natural signification relations are *used* in a specific way, to indicate *this* rather than some of the other states such a sign might naturally be correlated with.

That is, Dretske states that what a Type II system represents is not what its elements indicate, but what these elements have the *assigned function* of indicating. Hence, a Type II-system represents whatever a particular element (symbol or sign) indicates in a particular context, in the light of the lawful and persistent correlations that link the sign to what it signifies, but excluding a variety of additional properties or states that the sign might also indicate, which are nonetheless not taken to be included in the set of properties or states the sign has the explicit assigned function of indicating.

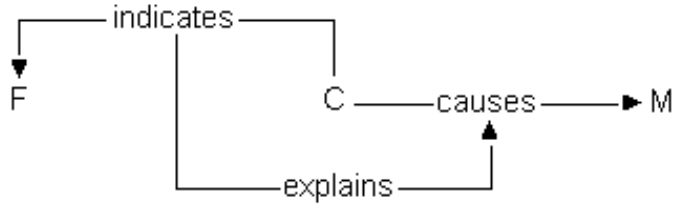
For instance, as Dretske notes, the functioning of a fuel gauge might be influenced by a variety of additional states and processes other than the

amount of fuel in the tank (gravity; electrical phenomena in the wiring; the tilt of the tank and/or the car [if the fuel gauging mechanism utilises a floater to measure the height of the fluid level], and so on), and as such instantiate an indication of these additional phenomena, but that is not what the gauge is intended or supposed or determined (by the designers and users) to do. The gauge does *indicate* these additional phenomena, but does not *represent* them.

This means that, as opposed to a Type I system, a Type II system is constrained by its role as a natural sign to indicate what it does. That is, because of its specific causal structure, there is a fact of the matter concerning what it can do, hence limiting the ways in which we might interpret what is represented in the system. But, given the constraints of these natural signification relations, Dretske says we can assign specific indicator functions to such systems as we construct them.

(Natural) Type III Representation: Type III systems, *natural systems of representation*, have the function of indicating something intrinsically, by virtue of the fact that the system's elements developed (evolved) with that explicit function within the system itself. These elements are natural signs - dark clouds approaching indicate the increased likelihood of rain falling here soon, and no interpretation is needed for that to be true. Despite the fact both Type II and Type III systems utilise natural signs, there is an important difference: the indicator function of Type II representation is *assigned*, though constrained by the causal specifics of the natural signification relation that is utilised. The function of a Type III representational system can only be *discovered*, it being what it is by dint of its evolutionary heritage, hence independent from any interpretative action we might undertake.

Dretske couches his discussion of the properties of various kinds of representations in an overarching question about the causal efficacy of internal states (a reason or intention, say). He uses the following schema to explain his views:

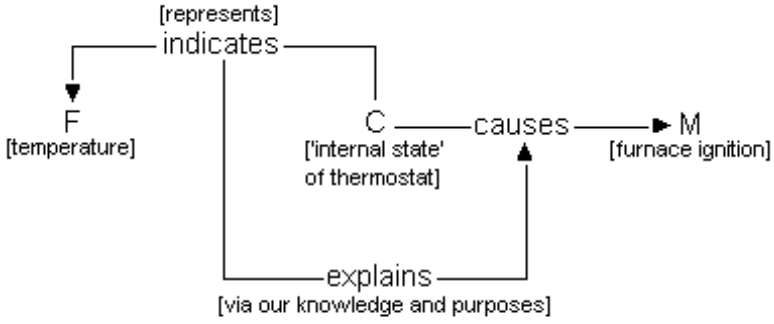


[Figure 17: representation, adapted from Dretske (1988)]

With this schema, Dretske means to elucidate his claim that the indication relationship between internal state C and external state F (i.e. C indicates F) explains the causal efficacy of C in realizing the occurrence of M.

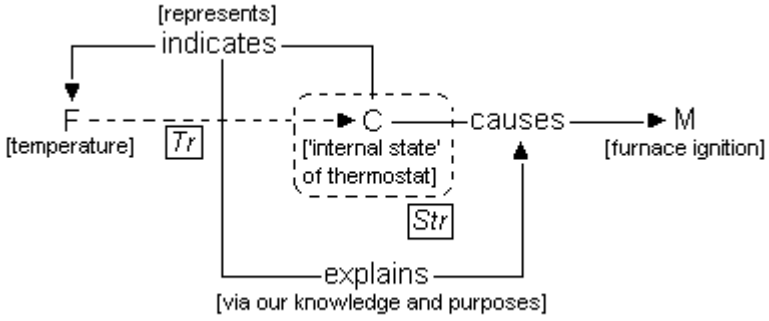
One of his famous examples concerns a thermostat. Central to such a device is a strip of bi-metal, which actually consists of two strips, each of a

different kind of metal, attached to eachother. Because the different kinds of metal expand and contract in different ways when the temperature rises and drops, the bimetal strip bends one way or the other during a change in temperature. Now, the thermostat has been built in such a way that as the temperature drops, the strip bends towards an electrical contact, closing the circuit, thus igniting the furnace. As the temperature in the room rises, the strip bends back, eventually breaking the circuit, thus causing the furnace to cease operations once more.



[Figure 18: representation in a thermostat, adapted from Dretske (1988)]

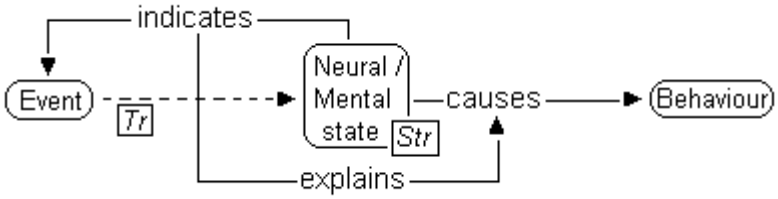
How does the system's internal state (the strip of bimetal being in a certain position) cause a change in temperature? There are two different kinds of causality involved: the drop in room temperature is a *triggering cause* (designated with 'Tr' in figure 19) for the thermostat to start to function, whereas the fact that the thermostat was designed and built to perform the function that it does in a particular way is called a *structuring cause* (designated with 'Str' in figure 19):



[Figure 19, representation in a thermostat, causal structure]

For a scenario involving a mental state, the structuring cause involved refers to the mental state being part of a system that has evolved to function

in a particular way in its environment, and the causal efficacy of the mental state is 'triggered' by virtue of its indicating some external state or event:



[Figure 20: mental representation]

In addition, the properties ('structure') of the mental state also contribute to the explanation of the resulting behaviour, because the behaviour is triggered by that mental state, in virtue of it being triggered by the external event it indicates.

Now, the behaviour of the strip of bi-metal can be understood as a *representation* of the temperature in the room, but it is not a *symbolic* kind of representation that $E_{(A)}C$-supporters are opposed to so vehemently. The same story can be told about mental states: saying that a mental state is a representation does not equal saying that a mental state is a symbol (or constellation thereof) that performs a particular function in a language of thought. In other words: an explanation involving representations need not be a profoundly computational story.

This undercuts some of the most consistent dynamicist criticism on representational explanations, at least for representations of the non-symbolic kind described above. Prinz and Barsalou (2000) note how Van Gelder (1998, 1999) suggests that the inherently coupled nature of the components of many dynamical systems presents a relation that is much more subtle than representation. Prinz and Barsalou object, and I would tend to agree, that the kind of representation at work in the thermostat case exhibits exactly the kind of coupling Van Gelder should claim disallows representation-nvolving description: the thermostat and the room's temperature are *coupled dynamical systems*.

I submit it is possible to understand Dretske's indicative causal link in terms of embeddedness, hence, compatible with at least some forms of $E_{(S)}C$. That is, Dretske's story would allow us to formulate an explanation of the causal efficacy of some internal state C of an embedded agent by virtue of the indication relationship between C and F obtaining, and defining that interrelatedness (i.e. embeddedness), or at least the ability of the agent to engage in such interrelations and no less than one prior actual occurrence of such an interrelatedness, to be an essential precondition for C being a mental state.

I make no claims as to Dretske's acceptance of my extrapolations - I merely use his ideas as a jumping-off point for my own speculations. Some work still needs to be done to turn the story above, about internal representation in thermostats, into an adequate account of the kind of representation that would contribute to cognitive processes, especially if we wish to understand cognition in $E_{(i)}C$-terms. Now, the idea that I should want this is perhaps a bit odd: is it not true that much of the work done in the field of $E_{(i)}C$, and of $E_{(A)}C$ in particular, is geared towards *abolishing* (the need to speak of) internal representations?

In the case of mental representation (rather than the internal representation of thermostats), I would hypothesize that an internal state would acquire Type II representational status by virtue of the fact that the system of which that state is an integral part has acquired, through evolution, a specific meaningful relationship to (certain features of) its environment - the system in question comprising the embodied agent and the structure of physical and social affordances he is embedded in. This description as such does not distinguish it from Type III representation (see below); the additional feature of having a function *assigned* to a naturally occurring correlative state that appears vital to Type II representation might, in the case of mental representation, be understood in terms of the multiple realizability of goal-directed behaviour. That is, an important aspect of having a mind is being able to autonomously pare down the extant degrees of freedom, i.e. design a creative yet effective act in a complex, dynamic environment of shifting conditions, based on the occurrence of a particular Type II representation. The idea is that this paring down need not follow a predetermined scenario: there is, in some sense, freedom for the system to 'pick' any one of a number of different possibilities.

However, this can constitute only *part* of the assignment feature essential to Type II representations: it captures the idea that one mode of realisation is arrived upon from a number of different options. However, it leaves open the matter of *who* or *what* is doing the assigning in this case. My suggestion is that care must be taken to avoid saying that the internal representative state as such is introspectively judged or interpreted by the agent, to be subsequently used as the basis for some act - this is exactly the kind of view I've been trying to get away from, in accordance with one of the most strongly held $E_{(A)}C$-views. Luckily, there is another option available.

I can sketch some of the particulars of this alternative by revisiting the example of colour vision. In the case of colour vision there is a particular constellation of internal states (the activations of different areas of the visual cortex) that stand in some causal relationship to features of some external object, such that there are sufficient grounds to call the co-occurrence of certain features of the object (microphysical surface properties) and activation patterns of the visual cortex a correlation of the kind required for the latter to be an internal representation of the former. Obviously, metamerism (see note 20) and the stochastics involved in the realisation of neural states disallow a perpetually valid one-to-one correspondence

between these two states, but the causal link obtaining, however multiply realizable it is, suffices here.

Now, this neural activation plays an essential role in causing, yielding or instantiating the phenomenal state associated with perceiving the colour the external object is identified as having. Decades of discussion about the status and properties of qualia tells us that what the exact relationship is between the neural state and the phenomenal state is as yet undecided, and suggested solutions to the problem are always controversial, but I will suppose it to be relatively *un*controversial that the former (the neural state) has at least *some* role to play in the occurrence of the latter (the phenomenal state). This is enough: for it does not make a difference whether there is a direct stimulus-response-reaction ummediated by phenomenal 'feels', or whether colour qualia perform some kind of motivating role, as long as the fact that there is that particular kind of neural activation that actually does *something* in the complex causal dynamic that makes the agent react to a colour in his environment. The neural activation need not even be the sole cause, or some kind of causal bottleneck that excludes other causal relations; if I were to hazard a guess, I would even say the real causal story is not nearly so simple, involving complicated interlockings of dynamical processes on multiple spatial and temporal scales. It is up to the biologists, neurophysiologists, computational neuroscientists and other experts to provide a more detailed account of the exact neural processes involved, but I believe it is fairly evident that without the activation of the visual cortex, without the activity in that brain region being correlated in a structured fashion with the external stimuli, much or all of what we would consider colour-related behaviour, at least the more complex and 'thoughtful' behaviour, would be impossible.

Now, that this causal link obtains and that an internal correlative state evolved into a state that has a particular role to play, makes this internal state a Type III representational state. The fact that the agent himself has the capacity to pick one of several different possible behavioural responses to coloured objects in his environment, and that this Type III representational state plays a role in that agent-environment interaction-dynamic, yields an interesting conclusion. The assignment feature to transform this Type III representational state into a Type II representational state is the paring down of degrees of freedom for action *by the agent himself*: with his actions, as embedded in a particular environment, the agent imposes a specific function on that internal Type III representational state. It is the embodied and embedded action that makes the difference: the evolutionary assignment of a function to a particular correlative structure can, on its own, merely lock in a Type III status, but the agentive action related to that Type III representation turns it into a Type II representation.

What is important - nay, positively *crucial* - to realise is that while these internal representational processes play a role in generating cognition, *they are not cognitive themselves*. Cognition is understood as something that an embodied agent as a whole, and as embedded in a particular physical and

social context, does: it is a process at the personal level. These *subpersonal* representational states *contribute* towards a complex dynamic of interlocking processes at different spatial and temporal scales (as hypothesized a few paragraphs ago), but, and I guess this is (part of) the problem, cannot be invoked to provide a complete explanation of cognitive processes. We certainly need them - they tell *part* of the story -, but there is more to be said, even when we were to obtain a perfect theory about the workings of the brain.

There are two reasons for this incompleteness, the first being that the vernacular we use to explain cognition in everyday circumstances, namely *folk psychology*, is formulated in terms of agentive behaviour and social interaction, not in terms of neural processing. The second reason is that, in most cases, it makes little sense to attribute sole causal responsibility to a neural region engaged in a particular activation pattern. This is because these neural activation patterns are themselves embedded, spatially (embedded in a constellation of other brain regions, encased in the head of an embodied agent that is embedded in and interacts with a complex, dynamic environment) as well as temporally (this dynamical neural system has a history).

This account is different from what Dretske (1988, pg. 88, 99) himself suggests. He claims that internal states, via the development of the organism they are a part of, acquire control of (say) the animal's limb movement. These internal structures, in being responsible for that peripheral movement, acquire an indicator function: because they indicate what they do, carry the information about some external state the way that they do, they exert control on the animal's movement. This indication relationship can also be incorrect, hence can *misrepresent* some external state: actions, as they are controlled by these internal states, might not always be successful.

The first modification that I wish to suggest is that internal states do not control external movements, but that they *contribute* to the dynamic agent-environment interaction (and contribute thus because of the specific correlation relation they exemplify with external states) which also includes contributions by bodily and environmental forces and inhibitions (i.e. enablings and constraints!).

In discussing Type III representational systems, Dretske says:

‘Can there be a serious question about whether, in the same sense in which it is the heart’s function to pump the blood, it is, say, the task or function of the noctuid moth’s auditory system to detect the whereabouts and movements of its arch-enemy, the bat?’ (Dretske 1988, pg. 63)

Perhaps there cannot be. But, to once again invoke the example of *colour* (see chapter 4), there can be, and is, disagreement about what the proper function of colour vision is. Obviously, something along the lines of 'to help

the animal interact with its environment' is correct, but saying this means making a fairly innocuous and non-committal claim. It says nothing about the underlying means and ends that are in play in colour vision: does the animal extract the surface spectral reflectance from incoming stimuli, or does it react to contextually meaningful invariants, or should we pick yet another description? This unclarity emerges because *what* colour vision does for the animal, *how* it does so and *why* it might do what it does appear to be different questions (this claim is in line with the differentiation between several notions of function as presented in section 4.5), and each question requires different strategies and supporting theories to receive an answer. How we describe its function is determined by context, i.e. how we conceptualise the colour-involving agent-environment interaction-dynamic, and at what granularity we do so - that *does not* mean we are free to assign a function (as we are free, to some extent, in the case of Type II systems). The function of colour vision within a particular context is still something that needs to be discovered, but which function we might discover depends on the direction of our investigation - the exploratory strategy we pick, and the intent with which we implement that strategy (i.e. the kinds of questions we feel compelled to ask).

Dretske couches such issues in terms of indeterminacy of function, with an associated *indeterminacy of representational content*. This is what that means: we assign to a fuel gauge the function of measuring the amount of gasoline in a tank, so the position of the needle on the meter represents the amount of gas. However, suppose the gauge does not distinguish between there being gas or water in the tank. Dretske says that if there is water in the tank, this does not mean the gauge misrepresents something; rather, it means that there is some indeterminacy concerning the fuel gauge's function. After all, it can be said to *correctly* represent the presence of a certain amount of *liquid* (rather than specifically 'gas').

We, as users of the fuel gauge, have the responsibility of checking whether the tank really does contain gas, and if it does, the gauge can perform its function as it was assigned by us. For systems of Types I and II, we as users have a lot of leeway, because the function of a system is what we say it is (obviously: given certain constraints, at least for Type II systems). For systems of Type III, determining what a system's function is, is shrouded in more uncertainty, at least to the extent that we need careful (scientific) investigation to discover that function: we have relatively little say in the matter. However, the story above, about embodied/embedded paring down of options helping to give a Type III representation certain Type II properties, muddles the issue once more.

The moral of the story is that the function of some representational system depends, to a large extent, on its context; this aligns quite nicely with the idea that the content of a concept depends on its context of use. For a weathered $E_{(i)}C$-supporter, and especially those inclined towards $E_{(S)}C$, neither of these conclusions is suprising: what cognition, or a formative

component of the cognitive dynamic, *is* or *does*, depends on the kinds of processes it partakes in and depends on.

As I have shown, Dretske suggests that the addition of mechanisms whose use depends on the representation correctly indicating something, can attenuate the uncertainty concerning what the function of a particular representational correlation is. If a representational system has the function of indicating the presence of a particular prey, and such a representation occurring causes the animal to perform a sequence of actions designed (evolved) to capture the prey, there is a criterium for the representational system performing the function it is supposed to in a successful manner: if it misrepresents, the action fails, and if this happens often enough, the animal dies.

So: this is my first modification of Dretske's idea: representational states of whatever type perform a function in a multi-layered agent-environment interaction dynamic, rather than being the sole cause of some action.

My second modification of Dretske's account is the point I already made above: representations (of Types II and III at least) *are not cognitive entities*. Dretske states (1988, pg. 99) that the process of learning confers a particular function, hence *meaning* onto the representational, indication-exemplifying structures. I would suggest, rather, that something having meaning can only occur at the *personal* level, not in the *subpersonal* realm. Meaning emerges in socially mediated, agentive interaction profiles: it requires some form of interpretation, and interpreting is what *persons* do, not neural regions or representational states. To scaffold this last suggestion, I will say more about the *emergence of meaning* in section 9.1. But for now I can say that these two modifications go to the very core of the RM-model, which is an attempt to provide a template for the description (and possibly explanation) of cognition, as it is to be understood within the $E_{(i)}C$-approach, and it is apparent that many different kinds of data, described in terms of the different spaces of the model, contribute to the total explanatory account.

Particular ways of utilising Dretske's theory could have implications that some $E_{(A)}C$-supporters would wish to avoid; Dan Hutto's insistence that he adheres to a bio-*semiotics* rather than a bio-*semantics* (see section 7.2) is an expression of such an inclination. The point I want to make is that a particular, qualified way of invoking representations of Types II and III in providing explanations *need not be* in any flagrant dismissal of the non-cognitivist groundrules laid down and adhered to by most $E_{(A)}C$-supporters. To strengthen this claim, I need to say more about the notion of *information* involved in these representational processes.

7.6 - Information

It is time to dive deeper into the complex concept 'information'. Note how Clark's (1997) general definition of representation (reproduced in section 7.4 above), invokes this notion:

"(. . .) let us call a processing story representationalist if it depicts whole systems of identifiable inner states (local or distributed) or processes (temporal sequences of such states) as having the function of bearing specific types of *information* about external or bodily states of affairs." (Clark 1997, pg. 147; italics not present in original text).

The most common accounts of information focus on syntaxis. Norbert Wiener (1961), for instance, one of the main inventors of cybernetics (the study and associated industry of transmitting and processing signals), states: 'Information is information, not matter or energy'. Expanding on this idea, Gregory Chaitin (1999) defends the thesis that information is muliply realisable. He writes:

"The conventional view is that matter is primary, and that information, if it exists, emerges from the matter. But what if information is primary and matter is the secondary phenomenon? After all, the same information can have many different material representations in biology, in physics, and in psychology: DNA, RNA; DVD's, videotapes; long-term memory, short-term memory, nerve impulses, hormones. The material representation is irrelevant, what counts is the information itself. The same software can run on many machines."

The sign might be understood as the basic component of information, or perhaps the carrier of information, or the entity in which information habitually congeals. Semiotics is the branch of science that studies signs, and is described in the on-line Merriam-Webster dictionary as follows:

"a general philosophical theory of signs and symbols that deals especially with their function in both artificially constructed and natural languages and comprises syntactics, semantics, and pragmatics." (entry 'semiotics', www.m-w.com)

These three aspects of signs or ways of studying them break down as follows:

(1) *syntactics*: this aspect involves the formal relations between signs. - coherence of sequence). A focus on this aspect supports multiple realizability of information most explicitly; still, there is always a substrate in which this informational structure needs to be realised.

(2) *semantics*: this involves the relations between signs and the world, i.e. how signs refer;

(3) *pragmatics*: this involves the function and evolution of signs, and specifies the relation between sign and user.

One could now ask what aspect of the sign actually carries the brunt of the signifying burden. It appears obvious that, if one accepts the essential role of contextuality (as is one of the main theses to emerge from SToCC - see chapter 6), the *pragmatist* dimension would have to be where much of the 'work' is done. That is, it is in the use of some decoding strategy applied to an information-bearing structure that the semantic and syntactic dimensions are attributed their content. The syntaxis will the main factor in constraining the probability for the (physical?) structure being picked as possessing an affordance of meaning-attribution, semantics emerges in the interaction of physical constraints (syntax) and utilisation (pragmatics). Hence, whether some syntactic structure actually contains information, and what kind of information it contains, is context-dependent.

Look at the way DNA can be said to contain information: it prescribes in a basic way what kinds of protein structures might or should be formed, but the way in which those are actually constructed, taking into account the many processes providing contextual influences (e.g. the parameters of the intra-cell chemical milieu, which can either inhibit or accelerate DNA formation), is *underconstrained* by the information actually present in DNA.

This would entail a kind of dispositionalism about information: information is characterised as a kind of disposition, i.e. as a property which plays a particular role in a particular context, in relation to a particular agent. Hence, it is possible to define a information-bearing structure as constituting an *affordance* (obviously: for some agent in a specific context). The impossibility of defining general and objective translation prescriptions for information-bearing structures and the associated inherent contextuality of information-use underline the role of some form of interpretation in the pick-up of information.

A general definition of information could then run as follows: information is a disposition to constrain probabilities of the unfolding of some causal process within a particular context, i.e. to nudge the system in question towards exhibiting structured behaviour. Or, more compact: information is a structural feature of some object, process or state-of-affairs that constrains and/or enables an *interpretational* process by an agent.

The suggestion I wish to make is the following: concept use, in its most general form, is information use (or at the very least involves a behavioural dynamics that is instigated by interaction with information-bearing structures), and information use can be described as interpretation. Now, it is important to realise that there is nothing that forces us to understand this process of 'interpretation' in line with the charicature of cognitivism, e.g. where even the lowliest mental process is described exclusively in terms of an introspective attending to the contents of one's experience.

Think of how a particular dance performance might be called an *interpretation* of a particular piece of music: the occurrent performance is based on or inspired by the song. Or, I submit we say the dance is *informed* by the music: the structure of the music imposes constraints upon, and discloses possibilities for (i.e. enables), the dancer's movements as they are understood as an appropriate expression of the structure and flow of the music. Such an interpretation does not necessarily require a cognitive appraisal of the music's rhythm and modulations - an agent can, as it were, resonate along with the music, without having to think about the movements that need to be made. The claim can be stronger still: when a dancer thinks too much about what to do next, the resonance breaks down. A more pedestrian example is the following: try walking down a staircase really quickly, consciously placing your feet on the steps. Chances are this is not as easy as it sounds, because the considered placement of feet interferes with the natural movement involved in letting gravity and leg joint dynamics do their work. Thinking too much in this way will result in a jerky, unnatural, possibly even unbalanced and dangerous hobbling down the staircase.

This uneasy interaction of higher, supposedly 'representational' thought, and basic bodily dynamics suggests the familiar gap between the two. This gap is what I would like to do away with; however, in Hutto's discussion of biosemiotics (see section 7.2) it still appears to be present. His account concerned a characterisation of basic-level interaction that does not involve semantics; it is at higher, more complex levels that somehow notions such as meaning, content, representation and information are supposed to acquire a stronger descriptive presence. After all, Hutto does not deny that content and the like are appropriate ascriptions at the higher, linguistically mediated level of agents in a social world.

Now, there are certain aspects that appear to show a certain degree of similarity at both low and more advanced levels: the agent and his appropriate environmental niche have co-evolved in such a way that the agent is predisposed towards reacting in a fitting way in certain situations: this constitutes a coupling of these two systems, in which properties and powers of one influence the other, and vice versa.

My suggestion is that the content-involving properties and regularities that are present at that higher level, are higher-order expressions of processes that occur at the more basic levels. This *does not mean* I suggest an unmitigated return to standard cognitivist descriptions of *subpersonal* processing involving content, representations and information; rather, I suggest that the relevance of these basic-level processes should be understood in *personal* terms, i.e. *the embodied agent as a whole, acting in (and interacting with) a particular physical and social environment*. Some of the higher-order agentive descriptions of properties and abilities connected to these various levels were given in section 6.9 and 6.10.

7.7 - Concepts and Content

The position about content that I wish to defend suggests a (possibly) unusual twist involving the idea of 'content' inherent in most theories of concepts. In general, mental content is defined as that aspect of a mental state that represents or refers to some other entity (or particular properties thereof). What that orthodoxy entails in a bit more detail is demonstrated by the following two quotes. First, Alex Byrne unpacks the notion 'content' in the following way:

'Contents are *propositions*: abstract objects that determine possible-worlds truth conditions. Three leading candidates for such abstract objects are Fregean Thoughts[NOTE 70], Russellian propositions (structured entities with objects and properties as constituents), and Lewisian/Stalnakerian propositions (sets of possible worlds).' (Byrne 2004; note not present in original text)

So, according to Byrne, representational content has a structure and purpose congruent with the structure and purpose of a (linguistic) proposition: a representation being a certain way entails this representation having a truth-value, dependent on how it relates to the object, process or state of affairs which it is intended to represent.

Gareth Evans circumscribes content, and contrasts it with nonconceptual content, as follows:

'In general, we may regard a perceptual experience as an informational state of the subject: it has a certain *content* - the world is represented a certain way - and hence it permits of a non-derivative classification as *true* or *false.* For an internal state to be so regarded, it must have appropriate connections with behaviour - it must have a certain motive force upon the actions of the subject... The informational states which a subject acquires through perception are *non-conceptual*, or *nonconceptualized.* Judgements *based upon* such states necessarily involve conceptualization.' (Evans 1982, pg. 226-227).

Hence, according to Evans, experience has a certain representational content, and this content can represent the world correctly or incorrectly; judgments about experiential content involve or are constitutive of concepts, and these judgments can be true or false. This means that an important role that content plays, by virtue of it representing properties of some entity, process or state of affairs, is the role of infusing the concept (which it is a content of) with the possibility of assuming a truth value - assessing the content of a concept as it is used allows a judgment about whether an agent (the purported concept-user) actually possesses a particular concept (or possesses the *correct* concept). Hence, by virtue of an agent's concept being or involving a contentful (mental) state or object, it should be possible to judge whether this concept correlates with the appropriate aspect of the

world (a particular state of affairs) in either a truthful or an erroneous fashion.

It is a small step from judgments about having the *correct* concept to judgments about having a particular concept *at all*: if someone has formed a particular concept of some state of affairs, and this conceptualization turns out to be incorrect, do we say that he has the concept anyway, but an incorrect version of the concept, or do we say that he fails the criteria of concept possession in this case? Peacocke (1992) for instance, notes that possessing a concept means meeting the concept's 'possession condition', comprising the kinds of inferences a person should be disposed towards for him to have a full mastery of the concept in question. If you do not meet the conditions, you do not have the concept in question.

If we accept the application of truth conditions to representational content *and* to the judgments of representational content that yields concepts, there is a definite temptation to attribute to content the role of affording a rather strict concept-world correlation, a concept being a kind of judgment or apprehension of experiential content (consisting of representations). And when the meter reads 'false' in both cases (the 'world-representation'- and 'representation-conceptualizing judgment'-correlations), to deny the agent in question the possession of a particular concept.

For SToCC/RM, the story runs differently. It is the prevalence of 'truth'-talk involving conceptual content that I wish to tone down significantly, with the SToCC/RM-model in hand. More specifically, I take issue with the 'truth'-talk in a *binary* fashion as applied to concept *possession*. Because the use of concepts is an essentially contextual affair - concept possession is socially mediated, expressed in terms of situational agreement amongst discussion partners (i.e. social affordances), and further constrained by environmental affordances - the notion 'truth' is hollowed out. That is, on SToCC/RM, concepts usually do not admit of clear-cut, black-and-white truth-values.

Recall the claims in section 6.11.5, where in distinguishing SToCC from Conceptual Role Semantics, one of the claims concerned the former's use of appropriateness-of-use-conditions rather than the truth-conditions espoused by the latter. This intuition is strengthened by the properties of the granularity operator, which enables two discussion partners to have the same concept at a certain granularity (that lies within the bounds specified by the low-detail similarity-criteria inherent in normal conversational exchange), without having *exactly* the same knowledge (down to the very last detail) concerning the concept's correlate (see section 6.8).

However, even with this modification, I would wish to claim that there cannot be a neat separation between correct and incorrect. SToCC suggests that after tallying up the scorecards for possession conditions, we are not left with binary values; rather, appropriateness-of-use-conditions admit of many intermediate values. An exacerbating factor in this case involves the multiple realization of concepts in SToCC: there are many

different ways of 'having' a specific concept, in part because the possible behavioural profiles associated with a concept usually range rather widely.

But the above appears to cause a problem for SToCC - or rather, increase the urgency of a problem that has been lurking under the surface for a while now. The problem is this: SToCC is a theory in the $E_{(i)}C$-tradition, hence is supposed to be careful about invoking 'representations' in its explanations. It should be clear that SToCC endorses the enactivist claim that basic sensorimotor interaction does not require *symbolic* representations: the scenarios in play at this level more closely resemble the interlocking dynamics of Millikan's intentional icons (see above, in section 7.2). And what was stated above amounts to the claim that even if there were such representations, talk of truth-conditions or ascription of concept possession in a binary fashion would be problematic: possession and correctness-of-use of concepts are to be judged in a gradual fashion.

Despite fuzziness invading the concept-world-correlation in SToCC at two fronts, I wish to maintain there is a role for content to play here. If, on a more mainstream understanding of concepts, a role that content plays is the role of infusing the concept with the possibility of assuming a truth value (as stated above), the task to be carried out now is to find an $E_{(i)}C$-compatible account that enables the existence of appropriateness-of-use-conditions.

In terms of the SToCC-account, 'having' a concept means being able to act/speak/think appropriately in a particular context (see section 6.3). In that sense, the role of content is to lock in what 'acting appropriately' means: in general, ascribing the appropriate content (i.e. at a reasonably accomplished granularity level) of the concept 'Eiffel Tower' to another agent means, amongst other things, that if one asks this agent where the Eiffel Tower is located, we expect him to say 'Paris'. And, in a less rigidly defined case, the content of the concept 'great white shark' will, to the vast majority of people, entail the imperative 'stay in the boat', rather than 'go skinny-dipping'. If someone chooses to do the latter, an appropriate response of ours can be to wonder whether this person has really managed to grasp the seriousness of the situation, has really understood what a shark is and is capable of - that is, whether he really does possess the concept 'shark' as we possess it.

The mechanism that realizes appropriateness-of-use-conditions for concepts to obtain, will have to do this by establishing an $E_{(i)}C$-alternative to the representational mapping that is invoked by the standard account: it is the representation that constitutes the connection of agent and world in such a way that an agent's decision to act in a particular way can be *right* or *wrong*, i.e. that there is a specific species of normativity involved. The issue of normativity will return later (in section 9.2); I will first attempt to flesh out the account of *content* that has so far been implicit in SToCC.

I submit that the RM-account already contains an $E_{(i)}C$-compatible mechanism that can perform the function that representational content

would on the standard theories: *the dynamics of constraints and enablings* (see section 3.5; more on this in chapter 8) constitutes a linkage between concepts on the one hand and body and world on the other. There is a specific account of *property realization* available which I will use to beef up the metaphysics of this dynamic.

7.8 - Dynamical Dimensioned Realization[NOTE 71]

In his (2002), Carl Gillett takes a stand against the theory of property realization defended by Jaegwon Kim (e.g. 1998). I do not intend to present to you a well-wrought position of my own in this particular debate, but I do believe that the account developed by Gillett can be of use in aiding my current argumentation. That is, his story about how microphysical entities can collectively yield new, non-reducible higher order properties has an interesting application in the kind of story that needs to be told about what kind of content there is in the SToCC-approach to concepts.

Gillett describes Kim's account of realization as *flat*: any and all properties or powers of an object can ultimately be described in terms of the properties or powers of the object's microphysical constituents, and/or constellations of such constituents and their resultant powers. Kim says: "It is evident that a second-order property and its realizers are at the same level (...) They are properties of the very same objects" (Kim 1998). This is a reductionist view, entailing that a complete explanation of causal interactions involving macrophysical entities ultimately boils down to a story about causal powers at the microphysical level.

As an alternative to this, the standard theory involving *flat realization*, Gillett offers his own account of *dimensioned realization*. Characterizing a property as something that can be individuated in terms of the powers it contributes to an individual, the idea is that, in macrophysical objects composed of microphysical entities, new properties can emerge that are not reducible to powers had by the microphysical realizers:

"Property/relation instance(s) F1-Fn realize an instance of a property G, in an individual *s*, *if and only if s* has powers that are individuative of an instance of G in virtue of the powers contributed by F1-Fn to *s* or *s*'s constituent(s), but not vice versa." (Gillett 2002)

An example often invoked by Gillett involves *cut diamonds*: the power to cause scratches in glass is a power of the diamond, not of the individual carbon atoms. The idea is that the diamond has new powers that are not in any way amongst the powers of the individuals that it is composed of (the carbon atoms), so that an explanation of the diamond's power to scratch glass is incomplete if only the causal powers of the microphysical constituents are invoked: a proper explanation requires reference to the diamond itself. That is, the causal power of the carbon atoms that are relevant here is the power to remain a certain distance from eachother under high pressure and/or temperature (i.e. maintaining a diamond's

characteristic atom grid pattern), and this power contributes to the hardness of the diamond, but it is *not the same power* as being able to scratch glass, which is a power only of the diamond as a whole. This suggestion edges away from Kim's view in attributing unique causal powers (hence properties) to compound entities (e.g. macrophysical objects) that involve, but are not exhaustively described in terms of the powers of the microphysical realizers.

Now, recall that on the standard account of content, representations and concepts, a version of which is exemplified in the citation of Evans above (section 7.7), a representation-involving interaction imparts an agent's mind with content, and a particular representational content implies specific truth conditions.

In the SToCC-model, the behavioural expression about which judgments pertaining to appropriateness-of-use-conditions can be made, is the inferred account. Being able to generate inferred accounts is the criterion for having concepts, and the ability to have concepts depends on being embodied and embedded in a specific way. It should be clear by now that if there is to be talk of 'representations' in this model, these are not static mapping relations.

Instead, the generation of conceptual structure in SToCC is a diachronic constraining/enabling dynamic, and I wish to suggest that this dynamic is that from which talk of content (which is expressed in terms of inferred accounts) is derived, and the presence of this content implies the applicability of appropriateness-of-use-conditions. That is, the property of an agent of having concepts imparts appropriateness-of-use-conditions, and this property is dynamically dimensionally realised by the properties of the physical and social environmental processes that the agent is immersed in, as well as the biomechanical properties of his own body[NOTE 72]. The interaction dynamics that these properties play a role in constitutes constraints and/or enablings for processes to be described in terms of the entire agent's concept-use-involving abilities, i.e. the concepts that he 'possesses'.

Because all these constraints and enablings are reciprocal (see section 8.5), hence the properties of the agent's conceptual dispositions thus realised impose constraints and evoke enablings for the other kinds of properties and processes in return, this realisation process can be understood as a dynamical coupling of agent and world which yields a new, higher-order property (i.e. involving an agent-world structure with normative aspects).

In a first analysis, there might be a problem with suggesting that this realised, higher-order content-establishing property is supposed to inhere in an entirely new *individual*, as Gillett's theory would. Because of the inherently temporary nature of the resultants of the constraint/enabling-dynamic, this would imply these content-establishing properties are to be understood as individuals, as *things* winking in and out of existence at a

staggering rate. This would appear to bring us right back at the kind of suggestion that Hutto spent so much effort at undercutting.

However, there is no need for alarm in this case. The very core of the *dynamic dimensioned realisation*-suggestion is that the constraint/enabling-dynamic realises a new dynamic, namely a (proto-)conceptual dynamic, i.e. the dynamics of conceptual space, with an important new property. This new property is characterised by the power to immerse the agent in *normative* structures: appropriateness-of-use-conditions are now in play, and these are the conditions that perform the modified roles that would, in standard theories, be performed by truth-conditions (see sections 6.3, 6.11.1 and 6.11.5). Because conceptual space constitutes a description of concept-involving dispositions of the agent as a whole, the individual to which this new, normativity-implicating property belongs is an individual in the truest sense of the word, namely *the agent himself*.

Hence, in this sense this case of *dimensioned* realisation in establishing the agent's conceptual dynamic affords the agent/individual a special status: the new property involving normative structures can only emerge when all realiser properties are present and in working order *in concert*, hence these properties all contribute to realise the norm-involving situation, but the normativity itself is not in play at the level of the realisers. That is, any and all judgments based on the normativity involved can only be applicable to the agent *as a whole*, and *as immersed in his environment in a specific way*. When I steal a CD from a store, judgments involving wrong or right, appropriate or inappropriate or wise or unwise are not applicable to the hand I used to grab the item, nor to my arm, my central nervous system, or even my brain or my body (as an object), but to *me*: *I* am the one performing the illegal act, *I* am the one to which these norms should apply.

However, it is important to realise that this special status (of being subject to norms) is not something only humans can enjoy: it applies to *all* concept-using creatures to a greater or lesser extent, and per my claims in section 6.4, the set to satisfy this criterion is significantly broader than the mere subset *Homo sapiens sapiens*. When a dog knocks over a precious vase, we do not hold it responsible in the same way that we would an adult human, but it is nonetheless quite plausible to get mad at the dog, and that we attempt to demonstrate to it that this destructive behaviour is not something we would like to see repeated. When a strong gust of wind knocks down another vase, we might get mad again, but not at the wind, or at least not in the same, culpability-implying fashion as we would at a dog or human - we might instead be cross with whomever left the door open, even though this person did not touch the vase. Many things differ between gusts of wind and humans, but I submit that the difference which I choose to summarise with the epithet 'has concepts', is of crucial importance in our attempts to explain the kinds of norms an agent is subject to, hence the differences in responsibility-ascription[NOTE 73].

Perhaps now Crane's 'aforism', criticized by Hutto (see above, in section 7.2), makes a bit more sense. Crane said that the content of a sensation, if non-propositional, is "what one would put into words, if one were to have the words into which to put it" (Crane 2003). At the basic end of the spectrum, we do not have such words, which is what makes Crane's remark so nebulous, but if we suppose that these 'things', the content of which we cannot put into words, fall in a continuous spectrum with the kinds of 'things' we *can* express linguistically, the relation between the traditionally dichotomous kinds of entities *nonconceptual content* and *conceptual content* becomes somewhat clearer: they are different conditional arrays (i.e. of different complexity, and regarded in different ways, e.g. phenomenally and linguistically) that are realized by the very same process, namely the constraining/enabling dynamic due to influences of physical, social and biomechanical processes.

Recall (from section 7.3) that traditionally, representations are conceived as context invariant, disembodied and static. Note how the kind of agent/world interrelatedness instantiated in *dynamical dimensioned realisation* (DDR) is the exact opposite on all three counts - my claim is that I can use DDR as the basis of a kind of representation that is congenial to $E_{(i)}C$-approaches, to scaffold the account of $E_{(i)}C$-appropriate concepts in the Radicality Manifold (RM), the expansion of SToCC to be described in the next section.

This locks in at least part of the metaphysics involved in the agent/world-interaction-dynamic as suggested by SToCC/RM. Now, it is likely that some, or perhaps many $E_{(i)}C$-supporters, enactivists and dynamicists in particular, will object to the flirtation with representation that is implicit in the account developed above. However, I believe, based on the considerations above, that there are representation-invoking explanations available which are compatible with dynamical dimensioned realisation, and manage to avoid the kinds of reifying symbolicism which the $E_{(i)}C$-approach is committed to evicting from philosophy and psychology.

My claim is that DDR, information and Hutto's use of biosemiotics all highlight aspects of one and the same process. What DDR does is explain the emergence of the twofold *content-like* (i.e. performing the function content would in standard theories) property of (1) making environmental information available to the agent, and (2) having this information-involving interrelatedness introduce normativity. In classical terms, such processes would be claimed to involve representation and information.

= =

[SUMMARY of chapter 7 AND PREVIEW]

So far in this book, especially in the form of Thelen et al.'s dynamical movement planning field as a description of basic cognition-involving behaviour, Thompson's theory of colour perception and the behaviour-based concept definition from chapter 6, the enactive ($E_{(A)}C$) approach has

been relatively important. On the other hand, the classical theories of concepts that I have discussed make elaborate use of internal representations, and these two views cannot easily be synchronised. In this chapter, I discussed Dan Hutto's Radical Enactivism to try and get a clearer notion of the role of representation in my theory of concepts. Radical Enactivism espouses the idea that there should be no representational content (to be understood as 'things' that are 'in the head') at all, at least not in the description of more basic forms of enaction.

My idea, in contrast, is that there are accounts of representation available which help us do the work of representations (i.e. establishing a meaningful link between the agent and his environment) without requiring the ontologically problematic use of reified internal mental processing. First, Andy Clark's notion 'representation-hungry problems' established that there are cognitive tasks which do sometimes require the re-presentation, in memory for instance, of objects that are not reliably present in the immediate environment. Based on a discussion of the various kinds of representation as distinguished by Fred Dretske, my suggestion was that we can have internal states which do not need to be representations of the classical kind (internal representations of external objects), but which are nonetheless representations in the sense that they contribute to the dynamic agent-environment interaction because of the specific correlation relation they exemplify with external states; this interaction also includes contributions by bodily and environmental forces and inhibitions (i.e. *enablings* and *constraints*). These 'representations' are not mental entities, however, but constraining or enabling factors in an agent-environment interaction-dynamic.

This is a dynamic of mutual constraints and enablings: the bodily properties of the agent, his social and physical environment in interaction collectively realise a specific behavioural profile with new properties, chief amongst which is the property of realizing concept-involving behaviour. I called the metaphysical structure that is involved *dynamical dimensioned realization*. The agent himself, as embedded in particular environment, has concepts. The fluidity involved in these dynamic interactions implies that concepts are not subject to rigid truth conditions, but *appropriateness-of-use* conditions

This interrelatedness of the agent's bodily properties, the social and physical properties of his environment and the concept-involving behaviour that results from this dynamic interplay can be expressed in a model which uses spaces, and this model will be explained in chapter 8. The dynamic movement planning field from chapter 3, the phenomenal colour space from chapters 4 and 5 and conceptual space from chapter 6 were all precursors to this final model, the 'Radicality Manifold'.

= =

[8 - The Radicality Manifold]

8.1 - A Constellation of Spaces

Now the time has come to cash in on all the preparatory work done so far, and offer the final sketch of the 'Radicality Manifold'-model (abbreviation: RM)[NOTE 74]. This is a framework that is intended to describe the complex interrelatedness of agent and world in a conceptually tractable fashion. The RM is intended as a combined extrapolation of the SToCC model and Thelen et al.'s (2001) dynamical movement planning field; the idea is to describe not just conceptual dispositions in terms of an abstract space, but all aspects relevant to the specification of a cognitive agent's interaction with the world. Towards this end, the integrated cognition-action-world interaction dynamic is split into separate 'spaces', each allowing its own type of descriptive strategies; at that point, the goal becomes to specify how those spaces are fundamentally intertwined.

Before I turn to a more detailed description of these spaces, it is important to note two things. One, the spaces that constitute the RM are *domains of description*: the RM is, first and foremost, a tool of description - a rather complex metaphor, if you will -, of relating different kinds of data (namely, environmental, social, biomechanical, conceptual and behavioural) to eachother. Second, several of these domains are only separable in an abstract and epistemic sense, not in a straightforwardly ontological fashion. The idea is that the agent's perception/cognition/action-dynamic with the environment results in a system that is fundamentally holistic, hence cannot - in actuality - be separated into different sections without the loss of essential information. The way of carving up the interaction dynamic that is exemplified in RM is an *explanatory* tool, a way of describing this infinitely complex dynamic with the scientific tools at our disposal. Therefore, being a model that strives to accomodate the kinds of data and description we already have at our disposal, the spaces-description of RM might actually come across as quite traditional: in addition to conceptual space (C-space), describing the interrelatedness of an agent's concepts, there is behavioural space (B-space), describing the behaviour of the agent (including locution as well as physical action), biomechanical space (M-space), describing the physical, biological (and so on) properties of the agent's body, physical affordance space (P-space), describing properties of the physical environment with which the agent interacts, and social affordance space (S-space), describing properties of the social environment with which the agent interacts.

The resulting model offers a new metaphor for $E_{(i)}C$, a framework with which, I claim, we can generate a structured conceptualisation of embodied, embedded (and so on) cognition that is more comprehensive and complete than, for instance, Thelen et al.'s (2001) model or Thompson's (1995) theory on colour perception, and in general more transparent about the interaction dynamics involved in cognitive behaviour. In other words, I suggest the RM offers us a suite of tools with which to pick apart instances

of the complex agent-environment interaction dynamics, with the intent of providing an explanation of what is going on.

The interrelatedness of the spaces of the RM-model is, in some sense, similar to the way in which an embodied agent interacts with *affordances* in his environment, which is why the next section is devoted to a closer look at this specific notion.

8.2 - Affordances

For RM, the term 'affordance', already mentioned in section 4.5, is a significant concept. Gibson (1979) uses this notion to express the properties of the environment (or objects therein), defined in terms of the *action possibilities* of a particular organism. For instance, a standard-sized doorway affords unimpeded passage to humans, but not to elephants, and thermals afford flying to birds, but not to octopuses.

Norman (1999) coins the modification 'perceived affordance', to take account of the fact that an object's user might not see all affordances of that object. In line with this remark, it is possible to see that the use to which an object is put by an agent depends not just on the physical capacities and dispositions he might, in principle, have, but also (rather crucially) on the capacities that he knows how to actualise. Gibson defines affordances as properties of objects, and these properties are defined relationally, i.e. in relation to the possibilities for action of the agent. However, some of these possibilities might never be realised, *if the agent does not know how to do so*.

Suppose a cargo plane loses a crate filled with tennis equipment while flying over the Amazon rain forest, and this crate is found by a local native who has never seen tennis balls or tennis rackets before. He will not be able to detect the affordances those balls and rackets have for us because he does not know what these objects are, and what they are for: he lacks the conceptual abilities to perceive and actualise the affordances for playing tennis (and all it entails) that these objects have. Certainly, he will be able to find some use for these items, but the possibility that he arrives at a form of behaviour that we would readily describe as 'playing tennis' need not be realized.

For RM, the role of the 'conceptual system' will, at least in part, be defined in embodied terms. This means that the native Amazonian in the above example might obtain the concept 'tennis racket' in limited form if he *observes* someone else playing tennis: in that case, he will have acquired some detached knowledge of the object, and this will suffice for limited concept possession, provided that this knowledge fits in with the concepts he already possesses (many of which are likely to be embodied or embodiment-based). However, he will acquire a much more complete and versatile concept if he, by chance or by following example, starts playing tennis himself. In this case, an important part of his concept 'tennis racket'

will consist of embodied knowledge, i.e. of having personal experience of how bouncing balls off a racket *feels*, and of being able to aim tennis balls without consciously guiding his arm through the required motions.

This conceptualisation of affordance involves a complex relation, in which the physical properties of the agent and the physical properties of the object both constrain the possibilities for some events to occur: given my size and strength, I can lift a pebble, but not a city bus. Furthermore, some events in the interaction of agent and object might be possible, but not very probable. For instance, I can sit on a tennis ball and throw a chair, but it is much more likely that I sit on a chair and throw the ball. In addition to these body-based constraints, the agent's conceptual system further constrains what events are probable, or at least what events are probable as events that are brought about *intentionally*. For example, when I am put inside the cockpit of an airplane in flight, the probability of me being able to execute a flawless landing is not very high, because there is much about an airplane cockpit that I do not understand.

The events that are mentioned above, the ones that are constrained by both physical properties of the object and physical properties of the agent, are *behavioural* events: in a sense, the relation between an object's affordance and an agent's physical properties are 'mediated' by behaviour. The 'walk-through-ability' of a door is a possibility for action, and is only 'activated' when an appropriately configured agent actually performs the appropriate action. The influence of the conceptual system, as described above, adds another dimension: some affordances are only relevant when they are *understood*. This is especially clear in the case of more complex phenomena such as language: the instructions of an art teacher contribute towards understanding the affordances of charcoal, paint, canvas and so on, so it is possible to say that the teacher's words afford the creation of a work of art by me, but this effect evaporates if I cannot understand what he says, for instance because I speak a different language.

These examples demonstrate, in line with Gibson's definition but probably a bit more radical, that an affordance cannot be merely a property of an object (or the environment): if we want to explain the agent's interaction with his environment, we need to take into account his behavioural patterns, the structure of his conceptual system and the properties and abilities of his body in addition to the properties of the environment. This structure of interacting objects, aspects and properties becomes even more complex when another crucial element is added: the constellation of *meaningful* properties that comes into play when the conceptual apprehension of the world is involved, especially when this world includes *other agents* with their own conceptual systems and behavioural patterns. That is, the inclusion of the conceptual system causes properties and events to *mean* something to an agent, especially when the environment's affordances are defined to include the actions of other agents[NOTE 75]. This means that the behaviour of other agents is what establishes (part of) the world's affordance structure, alongside that environment's physical features.

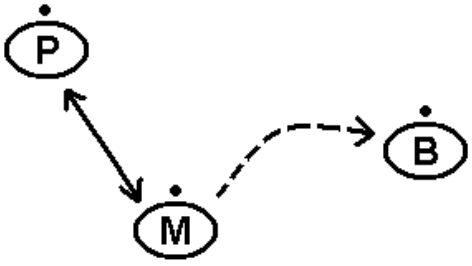
The Radicality Manifold (RM) is a model that includes a description of exactly these aspects of the agent-world interaction dynamic. Each of these aspects allows for a description in terms of an abstract space, that is similar to the account of concepts that was presented for concepts in the previous section (namely SToCC).

8.3 - Description of the Spaces

The RM portrays the interrelatedness of several domains of description, linked by recursive constraints on, or 'enablings' of, degrees of freedom between spaces, as well as explanatory connections. What this means exactly will be explained in the pages to follow, but one point needs to be made clear right now: these spaces are not located anywhere, least of all inside an agent's head; also, these spaces do not necessarily encode, exemplify, constitute or contain representations. What the RM does is offer a model to parametrize and describe the *behavioural dispositions* (including action, cognition and locution) of an embodied agent that is embedded in a particular environment.

Recall the movement planning field devised by Thelen et al. (2001), which I discussed in section 3.2; this model forms an important source of inspiration for my model in two ways. First, their model expresses a prioritarisation of certain kinds of data, which I intend to expand upon in order to provide a more comprehensive description of cognitive behaviour. Second, I will use this movement planning field as the template for behavioural space.

The general structure of Thelen et al.'s model can be depicted as follows:



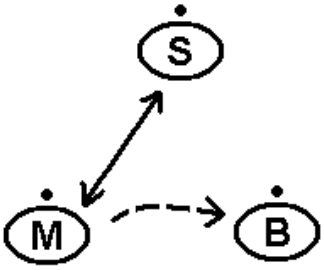
[Figure 21: interrelatedness of agent and *physical* world]

I have used this picture before: it also describes, in principle, the kinds of interactions that were conceptualised in the discussion involving ecological theories of colour perception (see chapter 4): properties of the environment (in the 'ecological colour'-case: structural regularities in the optic array) and properties of the observer (for colour: an agent with a specific retinal chromatic dimensionality, and biomechanical properties which allow him to perform certain actions in response to environmental structures) collectively determine the kinds of behavioural profiles this agent is capable of exhibiting (e.g. approaching food-shaped objects with a particular colour,

but avoiding objects of the same shape but with a different colour, such as ripe and unripe fruits, respectively).

That is, the biomechanical properties of the agent as they change over time (dM/dt), in *interaction* with physical properties of the environment as they change over time (dP/dt), can, at some level of detail, be *described* as behaviour (i.e. the change of behavioural patterns over time, dB/dt). This behavioural description can be expressed in terms of a behavioural space, as a higher-dimensional version of the movement planning field utilised by Thelen et al. (2001).

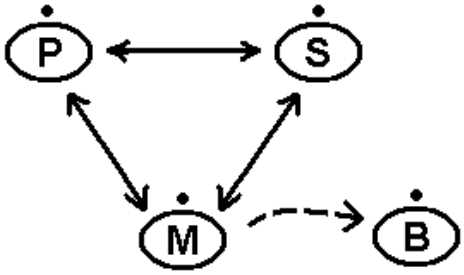
Mutatis mutandem, the structure underlying the discussing regarding the linguistic categorization of perceptual colour space (see chapter 4) might be characterised as follows:



[Figure 22: interrelatedness of agent and *social* world]

That is, specific socio-cultural properties (i.e. linguistic regularities, social customs) and the properties of an agent's perceptual system *also* collectively exert some determining influence on the kinds of behavioural profiles an agent is capable of exhibiting (e.g. implementing particular categorizations, such as using the same colour word in describing objects that are alike in the appropriate sense).

In reality, true $E_{(i)}C$ partakes in *both* types of interaction structure, for both physical and social environmental cues, together with the relevant properties of the agent, help constrain the kinds of behaviour an agent is likely to engage in. In terms of the depictions used above, this would look something like this:



[Figure 23: interrelatedness of agent with physical *and* social world]

So, the interaction of physical (P) and social (S) aspects of the environment plus the biomechanical (M) properties of the agent *collectively* yield a particular agent-environment interaction dynamic, of which we can describe part (namely, the motion of the agent's body) in terms of behavioural (B) patterns. As said, the model by Thelen et al. (2001) provides a DST-based description of this behaviour (B), i.e. using a movement planning field (see section 3.2).

In the following sections, I will specify the properties of the four (sub-) spaces depicted above, as well as the relations between them, in a bit more detail: Behavioural space (B-space), bioMechanical Space (M-space), Physical affordance Space (P-space) and Social affordance Space (S-space). Then I will specify how Conceptual space (C-space) fits into this structure, in light of the considerations about concepts in chapter 6, and the dynamical dimensioned realization of concept-involving behaviour as discussed in chapter 7. The specification of B-space, the first space to be discussed here, offers a general format in terms of which the other spaces can be modeled.

8.3.1 - Behavioural Space

As noted, Behavioural space (or 'B-space') is inspired by the dynamic field used by Thelen et al. (2001) to model an infant's movement planning (described in section 3.2), and my adaptation of their model also serves as an example for the characterization of the other spaces of the RM. Irrespective of the shortcomings I believe Thelen et al.'s theory to have as an account of cognition *as a whole* (see also section 3.2, and Van Leeuwen, 2005), their ideas have significant merit as a template for the partial (namely, behaviour-based) description of cognition-involving processes, and that is how their ideas will be used in what follows.

Recall that in SToCC, conceptual space was defined as a *dispositional space*, i.e. an abstract space describing what an agent might do (say, feel, think, ...), in terms of the concepts associated with those actions, in a particular situation. Each of the other spaces that constitutes the RM is also a dispositional space, a higher-dimensional version of a dispositional field, similar in form to the movement planning field used by Thelen et al. (2001). As described earlier, Thelen et al.'s model involves a dynamic function specifying a field which depicts a reach towards either location A or location B if it spikes beyond a particular threshold value. The dynamics of the field is determined by saliency of input (e.g. a brightly coloured toy vs. a bland-looking object), 'memory' of earlier tasks, and the cooperativity of the field (to help generate a single response with complex input, regions of the field that lie close together are mutually stimulatory, while inhibiting more distant activation). Thelen and colleagues used this abstract construct to model behavioural performance, and the predictions of the model turned out to align quite well with the behaviour observed in real infants.

It should be possible, at least in principle, to increase the dimensionality of this field to include all types of behaviour a particular agent is capable of. That is, it should be possible to specify additional axes in Thelen et al.'s very simple behavioural space along which to depict other parameters and variables, which would allows us to model not merely reaches to either location A or B, but also the strength and speed (and other aspects) involved in reaching, plus other forms of behaviour such as walking, running, head-turning, jumping, and so on. We can then attribute weights to specific combinations of activation levels, to model the likelihood of occurrence of a specific behavioural pattern for some agent in a particular context.

For instance, an agent with strong leg muscles and relatively low body mass is likely to be able to climb a staircase comparatively quickly; a tall person is much less likely to require a lot of exertion to reach objects placed on the top shelf in a kitchen geared towards regular-sized people; a very limber person is more likely to be able to execute certain movements than normal agents, and so on.

These examples demonstrate one particular point quite effectively: bodily motion-based behavioural patterns are necessarily related to environmental properties (P-space: the physical environmental affordances, for some agent, of a staircase, or a top shelf) and the agents' biomechanical properties (M-space: the kinds of dispositions and abilities that are possible due to having strong leg muscles, or being tall or limber). Speaking of other kinds of behaviour - cognition and locution - becomes possible when social environmental affordances (S-space) are introduced. These domains of properties interact, and the mutual constraints and enablings between these domains collectively determine the possibilities for action of the agent in his environment. It is these relations and *dispositional interactions* that are encoded in the spaces of the RM.

Just like having a concept was defined (in section 6.3) in terms of dispositions towards not only 'normal', physical action, but also cognition and locution, B-space should also contain cognitive and locutionary acts. In the case of cognition, this move safeguards against a slide towards behaviourism, which would make talk of the phenomenology underlying C-space, as well as much of C-space itself, inexplicable; in the case of locution, this is because speech is an exceedingly important aspect of our behaviour, and together with these other types of behavioural expression, it contributes to the shaping of the sociocultural environment for other agents.

8.3.2 - bioMechanical Space

Biomechanical space, or M-space, 'encodes' the properties of the body as they constrain or enable activities of the agent, and certain dispositional patterns in the spaces it is linked to (P-space and S-space). In other words, M-space describes the biomechanical and neurophysiological contributions or inhibitions to the actions of the agent as a whole, i.e. the suite of the

agent's physically specifiable dispositions as constrained by the properties of the cells, organs, joints, muscles and so on.

This space is the classic domain of embodied dynamics. Constraints are described as, for instance, possible movement dynamics in terms of properties of joints and ligaments and muscle strength, thus enabling particular behaviouro-motor patterns but disabling others (my arm can bend *this* way, but not *that* way), or endurance as a function of the capacities of the heart and lungs; the properties of the retina and additional downstream neural apparatus afford the perception of certain objects belonging to a certain range of colours, textures, sizes, and so on, and render properties and signals outside that sensitivity range (say, ultraviolet radiation) inaccessible.

The constraints and enablings mentioned above are bottom-up in character. Top-down influences, where certain M-space dynamics influence an agent's biomechanical properties, can be e.g. activity that depletes energy (or restores it: eating), increases strength and endurance (exercise), or causes damage (careless behaviour resulting in a broken leg or arm).

Obviously, M-space is structured. It should be possible to specify this structure in a layered fashion; here is a crude example:

Visual distinction abilities
→ (are explained by) →
(functional) properties of retina and subsequent neural processing
→ (are explained by) →
properties of retinal and neural cells
→ (are explained by) →
microphysical processes

It is important to note a few things about this schema. First, the overall behaviour of M-space is described in terms of the biomechanical and neurophysiological contributions or inhibitions to the activity of the agent as a whole, i.e. the suite of the agent's physically specifiable dispositions as constrained by the properties of the cells, organs, joints, muscles and so on. These contributions and inhibitions provide the parameters with which to specify the properties of the *contact layers* (see section 8.5 below), i.e. the description of the way in which the various spaces are related to eachother.

Second, the relations between the layers in the scheme given above are *explanatory*: a property at a particular layer is explained by properties or processes at some other layer. This implies that M-space also has a structure that can be defined in terms of a granularity gradient, just like conceptual space (see sections 6.7 and 6.8).

However, and this is the third point, we should shy away from attaching too much importance to the fact this scheme has a layered structure. That is, there is *not* a necessary endorsement about microphysical reduction to be

distilled from this schema. In RM, as in embodied/embedded theories in general, the agent level is ontologically most significant: when explaining macro-level behaviour (of whatever form), the biomechanical and neurophysiological properties and processes that usually matter most are structural assemblies that occur at that same macro-level, and the microphysical particles (whatever they are) are not necessarily the entities doing the relevant work. Explaining how these properties arise could obviously, in many cases, lead to explanations in terms of microphysical events, but I want RM to resist across-the-board reductionism, and granularity is the tool to accomodate descriptions of relevant processes at a particular 'level'.

This is important because macro-properties are not always exhaustively described as aggregates of micro-properties, where those micro-properties are ontologically primary; RM leaves room for explanations in terms of, for instance, self-organization and downward constraints and causation[NOTE 76], that could result in irreducible properties and structures. However, apart from the claim that the agent level, the 'top layer' of M-space, is most relevant, RM is largely agnostic regarding the exact specification of the structure of M-space, allowing a lot of room for explanations in terms of whatever theory (in the scheme given above about, for instance, the functioning of the visual cortex) currently works best. A particular array of constraints on C- and B-spaces can be derived from M-space descriptions in terms of biological, chemical, neurological, anatomical (etcetera) properties, but also in terms of higher-level dynamical processes. This also means that the RM-model allows for multiple realizability in terms of the material composition of the agent, but I would like to claim that it is likely that any agent that can be described in terms of the RM, and exhibits human-level cognition, will possess a human-like body. At the very least, we have not yet encountered or constructed anyone or anything that is capable of full-fledged human-level cognition, and is totally unlike humans in material composition and bodily structure. While this is by no means a decisive argument, it does contribute to arguments about the likelihood of some of our intuitions in this area.

8.3.3 - Physical affordance Space

Physical environmental affordance space (P-space) 'encodes' the environment's constraints on and enablings for the agent's behaviour (action, locution, cognition). The idea of 'affordances' was discussed in section 8.2, and there the point was made that affordances should not be thought of as merely physical properties of objects that are defined relationally depending on an agent's physical characteristics (e.g. a chair is of the right physical size, shape and structural integrity to afford sitting for normally abled, regular-sized humans), but should also accomodate the influence of the agent's knowledge and conceptual abilities.

It is important to note that all spaces that comprise the RM are perspectival descriptions of the entire agent-world-dynamic. Hence, this is also the case

of P-space, and this aligns quite neatly with what an affordance actually is: a specific possibility for action, emerging from the interaction of properties of agent and world.

Operationalising that interactive aspect, P-space can structured in terms of the *probability of interaction profiles*, an idea that I mentioned already in section 8.2. P-space too allows modeling in terms of a weighted dispositional space, and the biomechanical dynamics of the agent (i.e. the dynamics of M-space) co-determines the shifting array of probabilities in P-space. For instance, it is more likely that I sit on a chair and throw a ball than the other way around (unless I participate in a particular kind of talk show); both sitting and throwing are afforded to me by balls and chairs (as long as they are not too heavy), but not to the same extent.

8.3.4 - Social affordance Space

The affording dispositions of the environment to some agent modeled in P-space constrain and enable the dispositions in M-space, but also impart certain structures on an agent's *social* effectivities.
This means that if the depiction of the general structure of Thelen et al.'s (2001) model above, i.e. involving an interaction of physical constraints (P-space) and biomechanical properties (M-space) to yield a description in terms of behaviour (B-space) is correct, their model is incomplete - it lacks the social dimension (S-space).

We have already seen a very important class of socio-cultural influences on the ways in which an agent acts in his environment: chapter 4 contains an extensive description of the influence of language and customs on the behaviour of agents in response to specific chromatic stimuli.

Now, it does seem to be the case that constellations of P- and S-properties flow into eachother, and are difficult to separate in concrete cases. Whatever social affordances are created appear to be mediated, in a sense, by P-space: a social affordance always involves a physical object (an agent's body, or some arrangement or *design* of physical objects wrought by an agent, or even an acoustic phenomenon we interpret as meaningful speech) as that in relation to which we are supposed to exhibit a specific kind of behaviour.

Perhaps this would support the conflation of P- and S-space into a general affordance space (A-space), combining physical and social affordances, as these are all opportunities for behaviour, merely modulated by different kinds of external structures (objects as such for P-space, and objects arranged in a particular way, congruent with the intentions of another agent, for S-space). I do believe, however, that the addition of social structures results in a vast *qualitative transformation* of affordances: overlaying socio-cultural practices over a physical world introduces a specific kind of normativity to that world, introduces a highly context-dependent and fluidly dynamic suite of constraints and enablings to our environment, and moulds

the array of appropriate behavioural responses in ways that adhere to its own set of laws and regularities that are vastly *underdetermined* by the properties of the physical substrate.

One example of this confluence of P- and S-properties is the emergence of *meaningful* structures. The array of environmental affordances is constituted not only by physical affordances, but also by *semantic affordances*, and I assume that in a narrow sense, this can still be understood as part of ('encodable in') P-space. This affordance type includes, for instance, *contextualised* objects: a toilet, placed in a particular context in a museum, affords different actions than that same toilet in a bathroom (and different actions yet again when it is part of a display in a store), and might actually *mean* something to some agents, e.g. it could instantiate a particular statement by the artist, and spectators might (or might not) pick up on that.

However, and probably more significantly, semantic affordance also includes *other agents* as part of that dynamic environment - agents who say and do all kinds of things that are meaningful to the appropriately configured observer, and who engage in interactive behaviour to elicit equally meaningful behavioural responses from me. I suspect that the potential for profound, meaningful *interaction* is what sets this kind of affordance apart from the kind that is purely physically instantiated: a piece of text in a book and a personally delivered monologue by the writer might contain - in purely formal, logical terms - the exact same kind of semantic content, but the action opportunities afforded in these two cases are profoundly different. This results in the interesting situation where another agent's B-space contributes quite directly to my S-space (and vice versa): his behaviour constrains and enables my own actions and concepts in various ways.

This is why I have decided in favour of the inclusion of a separate S-space in the RM: despite the entangled nature of P- and S-properties, the kinds of dynamics I can - appropriately - exhibit in response is quite different.

There is more on the various kinds of influence (enablings and constraints) below: section 9.1 highlights the social, semantic relevance of physically instantiated patterns (what I call 'bodily syntax'), and section 8.5 describes some of the features of the constraints-and-enablings-dynamics that gouverns the interactions between the spaces of the RM. This latter description will be offered in terms of each space's separate contributions, inasmuch as it is possible to separate them - these will be abstract descriptions of properties and processes which, in actually, form a unified dynamic.

Important to note is that all varieties of semantic affordance are subject to *interpretation*, and in a context dependent way: a red light at a busy intersection is usually interpreted as demanding a different kind of response than that same shade of red in a fireworks display. And interpretation, of course, is bound up with the structure of C-space.

8.3.5 - Conceptual space

A lot has already been said about conceptual space (C-space), in chapter 6. There are two important steps left on the roster: explaining how C-space can be modeled in a way that takes B-space as its template, and integrating C-space into the RM-structure described in the sections above (i.e. involving B-, M-, P- and S-space). I will take these points in turn.

First, the modeling issue. Recall that having a concept of some object was defined as having the dispositions towards displaying behaviour (action, cognition, locution) of a specific kind in relation to that object. As such, C-space is a high-dimensional dispositional space *by definition*. Modeling these dispositions involves describing the concepts' enslavement hierarchies (see section 6.7) at particular granularities (see section 6.8). The fact that some aspect of a concept's internal structure (e.g. the enslaver) is more *prominent* than another lends itself to modeling in terms of dispositional fields with a particular weight distribution. This is how that works: in Thelen et al.'s (2001) model, the function that defines the field can be drawn up in such a way that a reach towards location A is much easier, occurs much more often, than a reach towards location B. Similarly, at some granularity, the bubonic plague's catalog of *symptoms* is likely to play a larger role in determining what the concept 'bubonic plague' means than some highly specialistic bit of knowledge about the *cause* of the disease.

Thus it should be possible to model some agent's C-space (probably merely in part, but theoretically as a whole) as an array of dispositions with particular weights, where those weights are determined, at least in part, by the constraints and enablings offered by, for instance, the agent's environment. That is, in a hospital's emergency room a properly embedded agent is likely to use emergency-room-related concepts (patient, trauma, medicine, nurse), in a way that depends on his role in that environment (is he a patient, a doctor or a concerned family member of a patient?), and also in a way that probably differs from the way said agent would be disposed towards using emergency-room-related concepts in, say, a shopping mall.

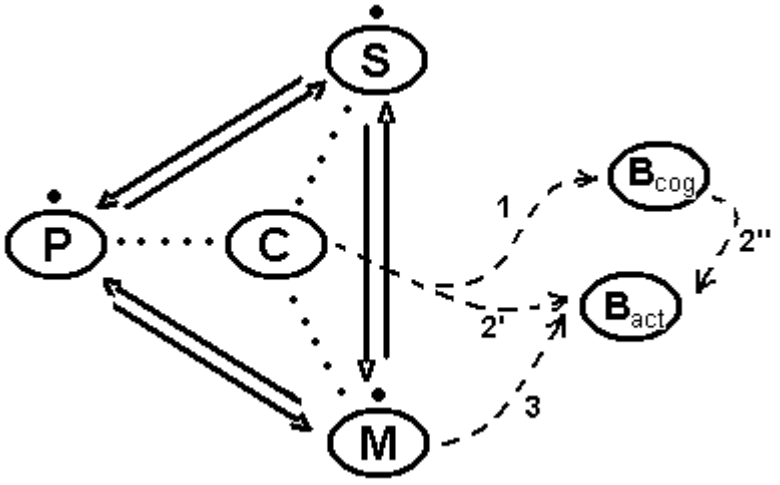
Second, there is the issue of integrating C-space into the RM-structure as it has been built up so far. In section 7.8, I outlined an account of the relatedness of concepts and conceptual abilities to the suite of bodily and environmental constraints and enablings, named *dynamical dimensioned realisation*. I submit that concepts allow an approach to behaviour that differs from what the model by Thelen et al. (2001) allows: *explanation*, rather than mere *description*.

In chapter 3, I posed the question whether Thelen et al.'s model was capable of doing more than merely providing an abstract, formal description of behaviour - whether it was capable of explaining that behaviour. The answer back then was: not very well. Conceptualising (pun intended) the agent-world interaction dynamic (the dynamic tangle of constraints and enablings of P-, S- and M-space) as realising a suite of conceptual

dispositions (which can be described in terms of C-space) *is* capable of providing explanations, I wish to claim.

8.4 - The 'Radicality Manifold'-Model

This results in the following, final rendition of the Radicality Manifold:



[Figure 24: **The Radicality Manifold**]

Legend:
⟸⟹ : mutual constraints and enablings
• • • • • : dynamical dimensioned realisation
`- - -´ : descriptive and explanatory relationships
1: ascription; 2': explanation (of basic bodily acts);
2'': explanation (of cognition-involving acts); 3: description

This is what I intend to depict here: conceptual abilities emerge, via dynamical dimensioned realisation, from the interaction of physical environmental, social and biomechanical processes. The dynamics of an agent's biomechanical properties can still be described in terms of behavioural, bodily acts (B_{ACT}), but the availability of a conceptual description allows an *explanation* (rather than mere description) of a particular behavioural profile. That is, the structure of C-space - spanning, as it does, a suite of abilities from basic body-based reactions (e.g. the Neurophysiological Yield, the NPhY, mentioned in section 5.2) to theoretic apprehensions of objects and occurrences (scientific conceptualisations of, for instance, 'colour') - allows one to provide accounts of *why* (not just *how*) an agent exhibits behaviour of a certain kind: this is because the properties of physical environment, social environment and bodily biomechanics conspire in such a way to create a structural regularity, i.e. a conceptual disposition, which - and this is crucial for the possibility of explanations rather than descriptions - is closely tied to an environment in which *norms* emerge (as per section 7.8 above, and section 9.2 below).

The breadth of C-space structure enables two explanatory routes: one directly towards the B_{ACT}-description, for low-level embodied behaviour (the properties of the NPhY emerging in the interaction of - chiefly - P and M explain basic categorization-based behaviour, such as bees flocking to flowers with colour X rather than colour Y), and one via the ascription of cognitive behaviour (B_{COG}). To elaborate on the latter case: many folk-psychological accounts presuppose (or even require) the ascription of mental states; such ascriptions to an agent are *justified*, I claim, by that agent *exhibiting the appropriate kinds of conceptual abilities*. Section 9.2 below will contain a brief exploration of the important role of *normativity* as it is involved in such ascriptions.

Please note that this comment about the possibility of explanations (giving reasons) rather than descriptions is not necessarily about what kinds of properties reasons have, e.g. whether or not reasons (as mental entities) can be causes (of behaviour)[NOTE 77]. Rather, this is to stress that invoking concepts enables us to provide interpretations and explanations of the behaviour of others (and ourselves), rather than being stuck at the descriptive level (which is where, for instance, Thelen et al.'s model still resides). The point of the RM model is that the interaction of biomechanical, physical and social properties is such that certain behavioural regularities are realised (via dynamical dimensioned realization) which allow the attribution of concepts to agents. This is all still purely descriptive: invoking concepts allows us to *explain* in addition to merely *describe*, simply because the kinds of explanations we usually give of agentive behaviour are defined in terms of having such and such a concept.

The non-reducibility inherent to Dynamical Dimensioned Realization exists in the fact that the attribution of concepts to anything other than an agent is incomprehensible. Or, formulated the other way around: if we are correct, under informed scrutiny, in attributing the having of concepts to something (be it a human being or another kind of animal), it must be an agent. In accordance with earlier claims (see section 6.8: the role of the expert in defining the intension of a concept), a claim can be said to hold up under 'informed scrutiny' if it can be made coherent within at least one scientifically sound paradigm. However, in most day-to-day cases, the scrutiny we subject claims to is not nearly so 'informed', allowing two agents to agree they share a concept as long as their discussion of that concept remains at a coarse-grained level. As we will see in section 9.2, the attribution of concepts is an integral part of the game of attributing reasons for acting, and this is mostly an epistemic practice: we use it to help us understand other agents and their actions.

8.5 - Bundle Dynamics: Functional Clusters and Contact Layers

It has become clear by now, I think, that the spaces that constitute the RM are fundamentally intertwined. However, there is a bit more to be said about this. For instance, it is important to note that C-space, at its perceptual/non-

conceptual base, blends into M-space. That is, the dynamical dimensioned realisation yielding very rudimentary conceptual abilities (for primitive animals) does not result in a C-space that offers much over and above the already instantiated M-space dynamic.

When we revisit the colour perception example, we can see that certain structural aspects of visual perception and phenomenology can be explained with psychophysical models. For instance, the specification of relations between colour experiences that can be expressed in terms of the dimensions hue, brightness and saturation - the three-dimensionality of phenomenal colour space - can be linked to the properties of chromatic sensors on the retina and the neural processing of their output signals (see Hardin, 1988). However, we should note that such models do not necessarily tell the complete phenomenal story: however sophisticated the models expressing the above-mentioned coupling of C-space and M-space might become, they are not necessarily capable of telling the whole story about the so-called 'what-it-is-likeness' of experience (see Nagel, 1974). As I do not presume to be able to solve the hard problem of consciousness I will put this issue aside (as mentioned earlier in note 36), and express an agnosticism about the role of phenomenal consciousness in the RM model: the main task to be carried out in this section is to sketch the model itself.

However, these remarks do highlight an important aspect of the RM, one that has been mentioned in more or less explicit ways before, and will be addressed more thoroughly now: the component spaces of RM are not ontologically separate entities, they are *partial descriptions*, each from a very specific perspective, of what is in actuality an integrated agent-world-dynamic.

There is a kind of *bundle dynamics* in place: very often, there is a necessary connectedness or reciprocity of processes that can be described in terms of different spaces. Some subset of dispositions of, say, M-space might enable specific dispositional patterns of a connected space - say P-space -, whereas another subset of M-space-dispositions might enable a completely different set of dispositions in that same P-space. After all, the kinds of options for an agent with specific biomechanical (M-space) properties to 'activate' a particular disposition is constrained by the kinds of activities afforded by the environment, i.e. the properties of P-space - a stairwell affords climbing (but *only* if the stairs are not too close together or too far apart, given the dimensions of an agent's body, i.e. the length of his legs, and the height he can lift them), a level field does not. All relations between P-, S- and M-spaces in figure 24 are recursively constraining/affording like this. So, the spaces of the RM describe an event sequence of an agent-in-the-world *collectively*.

The above implies that it is possible to distinguish several special relations involving subregions of the complete RM. I call these two-space relations the *functional clusters*[NOTE 78]. Spelling out the kinds of constraining and/or enabling properties that come into play between the various spaces will

help us to understand a bit more about the interaction dynamics that governs the relations between the various spaces. That is, not all properties and processes that can be described in a particular space are relevant for that space's interaction with another space, and even if some processes and properties are relevant at a particular time, they might not be at other times.

So, for instance, for the behavioural disposition of walking, the biomechanical properties of relatively large-scale assemblies of bones, muscles, tendons and joints in the arms and legs are relevant (obviously, this is not a complete list of relevant M-space properties). These properties of M-space are less relevant for the behavioural disposition of speaking, where much smaller forces, not to mention different body parts, are involved. Obviously, there is also some overlap: in both cases, lung capacity plays an important role. But the differences remain - consider, for example, how perceptive acuity is important in both cases, but in different ways: where a blind person and a seeing person might exhibit little discernible behavioural differences while speaking, they are likely to display significantly different behavioural patterns while walking.

These similarities and differences determine which properties from a particular space impose constraints or create degrees of freedom in another space, at a particular time, and how these influences manifest themselves. The simultaneous constraints and enablings at these various 'levels' can be mapped in terms of a multi-tracked granularity-gradient in each of the connected spaces: a specific subregion at some granularity can express properties that constrain or enable certain other properties or processes that are described in terms of a specific subregion at some granularity in the connected space, even while other properties, to be described in terms of those same spaces but at different granularities or in different ways, are also relevant.

For instance, while walking, properties of muscles and tendons (M-space) influence the kind of gait that can be achieved (B-space); physical properties of the visual and auditory systems (M-space) determine whether or not a particular obstacle (P-space) is detected, with obvious behavioural consequences (B-space: will the agent step up onto the curb, or continue as if he were expecting the walking surface to be level, and trip?); properties of the heart and lungs (M-space) influence the agent's endurance (B-space), and so on.

The theoretical vocabulary developed above allows us to get some grip on the *dynamic coupling* of each pair of spaces. A formal account of dynamic coupling in this sense would require the specification of how the variables of one system act as the parameters of the coupled system, and vice versa. An adaptation of that idea in the current context (i.e. as pertaining to the RM) will be given below, and it requires the development of one additional notion: the *contact layer*. This is what I call the array of properties and

processes in a particular space, at some granularity, that constrains or enables an array of properties in another, connected space.

Contact layers always come in pairs (one in each of the connected spaces), but there may well be several pairs of contact layers 'active' as a description of the way two connected spaces constrain or enable eachother, if processes at multiple levels or domains are in play. The 'walking'-example above demonstrates this, as several different processes encoded in M-space (muscle activity, processes in the perceptual system, respiration, blood circulation and so on) contribute to specific behavioural patterns.

I propose that it is, in principle, possible to specify the properties of each pair of contact layers in terms of a coupled pair of dispositional spaces, of the kind described in Thelen et al. (2001). The parameters to define the contents and properties of all spaces are described in a common format, namely an agent-centered description: all the processes in the various spaces are understood in terms of their contribution towards the totality of the agent-environment-interaction, and that contribution consists in constraining or enabling certain other processes, to be described in terms of one or more of the other connected spaces.

Some suggestions about what kinds of constraints and evocations of degrees of freedom occur in the various 'space-interactions' are specified below; this is by no means to be understood as an exhaustive description, merely as a coarse-grained indication.

One brief remark before I provide the list, about notation: 'Pm' in the list below, for instance, describes the 'physical/environmental-to-biomechanical' mapping, involving the kinds of constraints and enablers/degrees of freedom generated by physical environmental affordances (P) as they are relevant to biomechanical properties and processes (m). 'Mp' describes the the relation in the opposite direction, namely the 'biomechanical-to-physical/environmental' mapping, which is the way that an agent's biomechanical properties (M), possibly defined in terms of effectivities (see note 26), influence physical environmental properties and processes (p).

So, listed below are the kinds of properties in terms of which the relevance of one space for another is to be specified - see once again figure 24 for a general overview, and where in the RM these relations are to be 'located'.

C<—>P+S+M: Semiogenetic Engine

Special mention needs to be made of the dynamics that is instantiated in the relation between C-space and the constellation of P-, S- and M-space - what I wish to call the *Semiogenetic Engine*[NOTE 79]. The claim is that C-space has a special role to play in the grasping of meanings - attributable to objects and processes, or other agents - in the environment. In section 7.8, I described the metaphysics of this relation as one of dynamical dimensioned realisation, which plays a role in creating interactions between

agent and environment infused with norms; perhaps now I can say a bit more about the actual character of this suite of constraints and enablings.

Cpsm: conceptual knowledge and abilities can curtail certain forms of the P/S/M-dynamics: it might be possible to understand this as a form of downward causation by the dynamically dimensionally realised property of exhibiting conceptual dispositions on the ecological substrate, i.e. the dynamics of the embodied agent interacting with his environment. That is, when a P/S/M-system exhibits such a dynamics that a specific suite of conceptual dispositions emerges, new affordances congruent with that suite of concepts are realised as well. A practical example: if you have an incorrect concept of some object, or no usable concept of it at all, the resultant dynamics (to be described in terms of some behavioural response, in B-space) is unlikely to result in sufficiently effective, useful interaction with said object. See the 'amazonian tennis player'-example in section 8.2: the physical affordances are present, but they are not 'activated' because the agent does not have the right concepts.

At the cognitive level (i.e. involving the cases in which a particular conceptual dispositional profile justifies the ascription of mental states, in terms of B_{COG}), an example of this relation involves changing one's (metaphorical) point of view. The acquisition of knowledge can change semantic affordances; for instance, the way in which an unfamiliar device is approached and understood can change dramatically once its proper way of use is explained. Or: the way we see or attempt to understand an abstract painting can snap into place, perhaps like a kind of gestalt switch, when we are told what it is intended to depict.

PSMc: embodied feedback from environment involves a kind of 'hands-on' experience which modulates concepts. When I attempt to pick up a cup of hot coffee for the very first time, I am likely to learn something of vital importance about containers filled with a dark, steaming liquid: they can be hot, and hot objects are to be handled with care. A cognition-involving example of this relation involves interpretational constraints in terms of object properties, or action or linguistic expression by other agents. For example, when someone asks me a particular question, this imposes constraints on the kinds of answers that are thought to be appropriate (by either of us). This array of constraints and enablings contains the Perceptuo-Cognitive Degrees of Freedom: perceptual constraints, for instance in terms of the boundaries of perceptual acuity, influence the kinds of concepts and agent can have. For instance, the properties of my retinal cells and the neural cells that process visual stimuli determine what subsection of the electromagnetic spectrum I am sensitive to, and in what way (e.g. specific stimuli, namely those corresponding to a give agent's focal colours, are perceptually more salient than others). Damage to particular brain regions - a *biomechanically specifiable* situation - can constrain cognitive activity: the agent will have a deficieny in a specific subsection of his conceptual abilities because the relevant parts of his body are not capable of processing certain kinds of stimuli. Furthermore,

something as basic as an agent's size can influence the kinds of concepts he can have: my concept of 'tree' would be markedly different if I were the size of an ant, or an *Argentinosaurus*.

S<—>M: Social Preconceptual Processing

Sm: this relation involves, for instance, the kind of direct influence (i.e. in a fashion that is *not* consciously mediated) of another person's facial gestures, bodily posture and voice intonation on one's own bodily responses. In the theory of Gallagher (2005) this process is central to social interaction, and the practice of attributing mental states to others[NOTE 80]. A case can be made for the claim that much of our understanding of others in conceptual terms depends on these processes. Proximal physiological processes involved in this relation (M-space), most famously the activity of mirror neurons (see note 55), constrain and enable certain behavioural responses, and awareness of one's own bodily processes might, in some cases, lead to an understanding of the situation that elicits a modification of the appropriate concepts.

Ms: This is one's own influence on the other (hence, on S-space), which involves cases in which unconscious gesturing and vocal inflection influence the other agent. This is an *indirect* and *conditional* influence: it remains to be seen how much of one's signals are picked up, in whatever fashion, by the other agents, hence the constraining and enabling effectivity of this signaling behaviour on the appropriate aspects of one's S-space (namely, those other agents) depends on the behaviour of those agents, and the vast array of influences that underlies it.

M<—>P: Affordance-Effectivity Balance

Mp: the Mp-relation encodes a set of basic constraints within which an agent is, in a sense, free to design behavioural profiles. A doorway of a specific size and an agent of certain dimensions will only be able to interact in a limited number of ways, and the 'behavioural implementation' (i.e. what kinds of action are executed), given these properties, determines which specific affordances are at play in a particular situation. M-space dynamics (describable in terms of B_{ACT} - bodily behaviour) can influence P-space, for instance if the agent realises a change in vantage. An example is actively scanning the optic array (see Gibson, 1979): I can learn something about a particular object if I walk around it, possibly discovering more of the affordance constellation said object represents for me. M-space dynamics (describable in terms of B_{ACT}) can also consist in an active intervention in the world to change physical environmental properties: a tree has a specific affordance array to me, but when I cut down the tree and make firewood, the object and its affordances have changed.

Pm: this comes closest to the standard definition of affordances. An object can have specific properties, for instance spatial ones (size, shape), which instantiate particular behavioural opportunities for me, and different

behavioural opportunities for my cats. Conversely, the kinds of behavioural constraints such properties entail likewise differ from agent to agent.

S<—>P: Distal Ecological Dynamics

Sp: this relation includes the influence of socio-culturally motivated (meaningful!) activity on the arrangement of the physical environment. For instance, another agent may have designed an artifact in such a way as to enforce a particular reaction by other users, i.e. to elicit a specific Affordance-Effectivity Balance in that user's relatedness to the environment (or an object therein): a pair of scissors is intended to cut paper, and was designed by its designer to do that well above all other uses that object might have. The Sp-relation includes the possibilities for other agents to use our shared environment as a substrate for their own socio-cultural expression: a piece of art, undeniably a part of the physical environment, can be imbued with a specific meaningful content by the artist.

Ps: However, this artist too is bound by the constraints of that environment: the degrees of freedom of the Sp-relation depend on the degrees of freedom inherent to the other agent's personal affordance-effectivity balance, i.e. the constraints and enablings, presented by the physical environment to that other agent. This is the Ps-relation.

The relations above involve the kinds of constraints and enablings that co-determine the agent-world cognition/perception/action-dynamics. The influence by one space on the other can occur as modifications of dispositional weights of the other space's field, and/or as the forced phase change of such a field to include new parameters and variables. The main idea is that there be a structural coupling between adjacent spaces, where a change in one yields new constraints and degrees of freedom to be imposed on all other adjacent spaces. *Collectively*, this constellation of spaces describes an agent as he acts in his environment.

Obviously, this structure of subspaces is an abstraction of the integrated mind/body/world-dynamic, but this way of superimposing an explanatorily pragmatic structure onto this diffuse dynamical structure yields a framework for different kinds of data to be presented and interpreted in a useful manner: the tendency for some particular process to occur is analysed in terms of its contribution to the functioning of the agent as a whole. Put differently: dispositions in each space are defined in terms of the way they affect adjacent spaces (constraints and 'enablings' of either space's activity on the other), and (more importantly) their contributions to the functioning of the RM in its entirety.

For instance, the disposition of my ears and the auditory processing regions of my brain to react accurately to auditory stimuli within a certain bandwidth of frequencies and energies, allows me (the agent as a whole) to hear certain sounds, while others are rendered inaudible to me; what I can and cannot hear determines what I can have knowledge of, hence what I can

have concepts of; this influences my behaviour (either directly - a loud noise makes me jump up instinctively -, or via the conceptuo-cognitive route - someone says something and I can react appropriately because I understand him); and finally my behaviour and changes in my faculties for understanding (in terms of the concepts I possess) collectively modify the affordance structure of the environment as I perceive it.

I can provide another brief example, one which I hope will show how RM can help us get some grip on an instance of a complex interaction dynamic involving higher cognition. I suppose a good theory of embodied and embedded cognition would need to be able to incorporate an appropriately characterized *socio-cultural* dimension, so this example concerns one of the basic constituents of socio-cultural interaction: two people talking to eachother, communicating face to face.

In this complex interactional process, an agent's conceptual knowledge informs behaviour (including speaking, gesturing, unconscious body language), his behaviour (locution, body language) instantiates a particular affordance array for the other to relate to, and biomechanical properties (including mirror neurons, which respond to gestures, voice inflection and so on) inform the formation of concepts via conscious and subconscious awareness (which is generated in the picking up of affordances). Here we can see that all (sub-)spaces have relevant work to do, not a single one can be left out of the explanatory account, and changes in any one of the (sub-)spaces affect the entire RM.

All the processes and occurrences mentioned in these brief descriptions do what they do - *are* what they *are* - exactly because they stand in the kinds of relations they stand in. A neurophysiological account of auditory processing, for instance, has only limited use if it is not explained in terms of its broader context, which necessarily includes behaviour, cognitive structure and environment, and the way these hang together. This explanation 'in terms of its broader context' requires an act of *interpretation*, of devising how typical neurophysiological data is to be presented in order to be relevant for an account of cognition.

= =

[SUMMARY of chapter 8 AND PREVIEW]

This chapter saw the introduction of the 'Radicality Manifold'-model (RM), which is a metaphorical depiction of the interrelatedness of bodily properties, sociocultural environmental properties and physical environmental properties, and how these collectively yield concept-involving behaviour. The model describes behavioural (B-space), biomechanical (M-space), physical affordance (P-space), social affordance (S-space) and conceptual (C-space) properties. Each of these domains can be expressed in the form of a dispositional space, and the interaction between these

spaces is affordance-based: each property type creates constraints and enablings for the development of another property type.

This model expresses how conceptual abilities emerge, via dynamical dimensioned realisation, from the interaction of physical environmental, social and biomechanical processes. The availability of a conceptual description allows an *explanation* (rather than mere description) of a particular behavioural profile. That is, the properties of physical environment, social environment and bodily biomechanics conspire in such a way to create a structural regularity, i.e. a conceptual disposition. There are several kinds of interactions between the various spaces: the Semiogenetic Engine, Social Preconceptual Processing, the Affordance-Effectivity Balance and Distal Ecological Dynamics.

In chapter 9, a few implications of this model will be discussed: the emergence of meaning and normativity, the inherent circular nature of the model, the capacity of RM to individuate concepts, and the epistemological view that it implies.

= =

[9 - Implications]

The RM-model as described in the previous chapter is capable of accommodating several $E_{(i)}C$-appropriate insights, for instance about embodiment (see section 9.1). The model also holds a number of implications that need to be made explicit, about normativity (already referred to in sections 7.8 and 8.4, and explored in more detail in section 9.2 below), impredicativity (section 9.3), concept individuation (section 9.4) and epistemology (section 9.5). Investigating these implications is the purpose of the current chapter.

9.1 - Bodily Syntax and Meaning

Gallagher (e.g. 2005) suggests that origins of (meaningful) speech lie in synchronized expression through bodily gestures *and* vocalisations. My speculation, congruent with Gallagher's suggestion and very naturally expressible in RM, is that this synchronisation demonstrates a multimodal metaphorical mapping ability that is present from a very young age. That is, 'meaning' (of expressions) emerges from shared behavioural structures that are, out of necessity, embedded in a meaning-containing interactive process, and this emergence can be described in terms of an increase in metaphorical mapping acuity (in the style of Lakoff and Johnson, 1999 - see section 6.9).

This how that works. In section 7.8, I suggested that concept-possessing behaviour emerges, via dynamical dimensioned realisation, from a particular structural coupling of agent and environment. I would like to put forth the ancillary hypothesis that the emergence of higher-order conceptual abilities from lower-level conceptual dispositions occurs in a transitional zone, commencing as soon as the infant is capable of social interaction (which is, to a limited but nontrivial extent, already in the womb), in which onto- and phylogenetically basic *bodily syntax* traverses a progressive solidification, towards the emergence of semantic content (or semantics-exemplifying action, if we care to steer away from attributing content). This idea, to be explained below, is in *direct opposition* to claims made by Gärdenfors (2000; see also section 10.2), who defends a theory which also uses conceptual spaces. Gärdenfors states that the semantic content of meaningful expressions is first generated internally, after which an appropriate syntactical structure is chosen with which to make said content public.

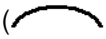
Rather, I would like to defend the idea that the exact opposite is true, both in the development of the child, and the evolution of our species' cognitive abilities (inasmuch as anything concrete - that is, anything beyond reasoned speculation - can be said about that development: the empirical base available to found phylogenetic claims on is notoriously slim): the structuredness of body-based interaction between agents (i.e. bodily *syntax*) comes first, and the *meaning* of behavioural, vocal and symbolic expressions emerges from those interactions. I suggest the hypothesis that

these behavioural structures are meaningful in part because they *evolved* as an inescapable (i.e. automatically occurring) mutual involvedness of conspecifics (e.g. mother and child).

Support for this notion, at least for its ontogenetic aspect, can be found in Stern (1985), when he speaks of *affect attunement*[NOTE 81]. The way in which an infant and its mother are able to share affective states depends, to a large extent, on mirrored behavioural structures.

Here, I will reproduce two examples from Stern (1985), to illustrate this phenomenon:

Example 1: "A nine-month-old boy bangs his hand on a soft toy, at first in some anger but gradually with pleasure, exuberance, and humor. He sets up a steady rhythm. Mother falls into his rhythm and says "*kaaaaa-bam, kaaaaa-bam,*" the "*bam*" falling on the stroke and the "*kaaaaa*" riding with the preparatory upswing and the suspenseful holding of his arm aloft before it falls."

Example 2: "A ten-month-old girl accomplishes an amusing routine with mother and then looks at her. The girl opens up her face (her mouth opens, her eyes widen, her eyebrows rise) and then closes it back, in a series of changes whose contour can be represented by a smooth arch (). Mother responds by intoning "Yeah," with a pitch line that rises and falls as the volume crescendos and decrescendos: "Y 

h." The mother's prosodic contour has matched the child's facial-kinetic contour."

The structure of these matched behaviour-aspects breaks down into the following more specific kinds of match (Stern 1985, pg. 146):

-*Absolute Intensity*: despite the difference in modality, the intensity of the mother's behaviour is the same as that of the infant's behaviour, e.g. the mother generates a loud vocal expression to accompany the infant's forceful arm motion.
-*Intensity Contour*: the intensity dynamics are matched, e.g. an increase and subsequent decrease in the infant's motion intensity over time is reflected in an increase and decrease of the mother's vocal behaviour.
-*Temporal Beat*: the matching of a recurrent behavioural component, e.g. a mother nodding her head in step with her child's arm motion.
-*Rhythm*: the matching of an irregular pattern of behavioural components.
-*Duration*: the behaviour of the mother and the child match in time.
-*Shape*: some spatial aspect of one agent's behaviour is matched in the other's behaviour of a different sort, e.g. a mother incorporating the vertical shape of her infant's hand motion by bobbing her head up and down.

Intensity matches (absolute intensity and intensity contour) occurred most often, timing matches (temporal beat and rhythm) somewhat less often, and shape matches (duration and shape proper) were least common.

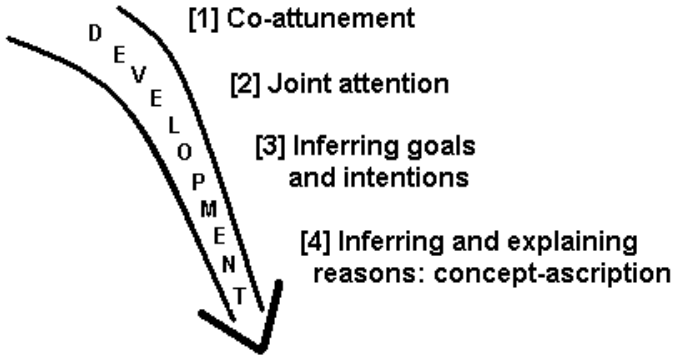
A highly significant feature of these interaction profiles involves their *cross-modality*: Stern reports the mode in which the mother reacted differed from the mode of her child's behaviour in 39% of the cases; in 48% of the cases, at least some aspects of the response-profile were different. For instance, in the first example above, the rhythm of the child's arm movements is matched by the rhythm of the mother's exclamations. In the second example, features of the girl's facial gesturing are matched by the mother's changing vocal pitch.

Because of their different modalities, these matchings are *not* imitations; Stern's suggestion is that these are matches between features of behaviour that express (some aspect of) the agents' feelings:

"Affect attunement, then, is the performance of behaviors that express the quality of feeling of a shared affect state without imitating the exact behavioral expression of the inner state." (Stern 1985, p. 142)

Of note is that, in the majority of cases (67%), the interacting agents are largely unaware that they are engaged in these matching activities: they are interacting, and the means by which they accomplish this are usually not controlled in a conscious fashion.

I suggest that we can view this interaction as the basis of what has been called *participatory sense-making* (De Jaegher and Di Paolo 2007), the social practice of collectively realising meaningful exchange.



[Figure 25: Development of participatory sense-making]

Figure 25 offers a rough sketch of how I view the developmental continuity realised throughout the social practice of (learning to engage in) participatory sense-making. [1] and [2] are very closely linked: often, we are engaged in processes of joint action, reaction and interaction. And this *social* interaction serves as the basis of 'stages' [3] and [4], yielding a complex dynamic of a kind of *negotiation in interaction* in attempting to achieve complementary and reciprocal goals, and at least part of this interaction requires joint attention.

This social co-construction of interaction structures can yield differentiations in the two-tiered normativity inherent to the environment that we interact with (containing affordances that impose physical and social norms upon agents: P- and S-space), when the social realm accrues a layered *deontic* structure. This stacked structure of socially co-constructed constraints and enablings contains individual desires which are modulated by communal norms (e.g. laws), which might in turn be held accountable in the light of independent normative entities (e.g. the rules of a language, which constrain and enable the development and upkeep of communal norms - this is a re-affirmation of the muted relevance of linguistic relativism, as I developed it in chapter 4).

My suggestion is that the ability to recognise intentions in one's social interaction partners, and the more advanced ability to to engage in an active *attribution of reasons for action* contributes to the ability to engage in *concept-ascription*. Concept-ascription involves positing or inferring predictions about *behaviour* (as opposed to positing the presence of *mental entities*). This constitutes a selective imposing of *norms*, i.e. making an assessment concerning the most likely or appropriate course of action for an agent, given the extant environmental and social normative structure. An agent's justification of his own concept-use involves explaining in what way his actions coincided with the observer's expectations, or explaining away a possible conflict between action and expectation, in some cases by providing arguments for the falsity of the concepts implicit in the oberver's demand for justification.

I do not believe this inferential process should involve an active, cognitive construal of the other's beliefs, reasons, conceptual dispositions and so on - often, the other's behaviour just *feels* right (or wrong). However, attempting to explain another's actions can involve such an overt construal strategy: concept-ascription can justify mental state ascription, at least as a folk-psychological working hypothesis, which in turn can serve to assist in *explaining* an agent's behaviour (see again figure 24).

9.2 - Normativity

In the above, I have attempted to clarify how the interactions of an agent with his environment involve an immersion in a two-tiered normative structure, involving physical environmental norms (to be expressed in P-space) and social norms (to be expressed in S-space). It is these norms which determine the *appropriateness* that has been present in the general definition of concepts that I have been using all along (see section 6.3):

having a concept A of some object/process/state of affairs O means being able to act in an appropriate manner, given the possibilities P for and constraints on action CA that O represents, and given additional contextual constraints CC.

This idea of our social practices in which we establish and explore physical and (especially) social norms shows clear parallels to the philosophy of Robert Brandom. Brandom (1994, 2000) defends a position he calls *inferentialism*, which constitutes a holist semantics that is generated in the social practice of *giving and asking for reasons*. This practice involves agents attributing commitments (being committed to playing the social game, with all it entails), acknowledging endorsements (accepting the behaviour of others as expressing a particular understanding of the world) and undertaking entitlements (underscoring one's own actions as being correct) (Bransen, 2000).

One of the motivating forces in giving and asking for reasons consists of embodied emotions (in the sense propagated by Damasio, 1998), and the phenomenology that goes along with them. Brandom shunts the effectivity of embodied emotions such as desires directly towards intentions: Bransen says of Brandom that he thinks...

"(...) desires are mere abstractions, they are what academically speaking would remain of intentions if one were to succeed in thinking away their role as practical commitments in inferentially structured action patterns." (Bransen 2000)

Now, 'intention' is an exceedingly slippery notion. My claim is that being $E_{(i)}$ means being immersed in an intentional structure in Brentano's sense:

"Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself..." (Brentano 1874)

That is, an agent is immersed, necessarily, in an environment, related to that environment and prodded to react by things in that environment. Hence, whatever internal representations of external states there are have this intentionality as 'aboutness' necessarily.

An intention in the sense of 'having the intention to act in such and such a fashion' is not the same thing, although it is a strongly correlated notion. I will call Brentano's 'aboutness'-intentionality 'intentionality$_A$', and the having of intentions to act a certain way 'intentionality$_B$'.

It can be argued that the former is a necessary precondition for the occurrence of the latter. That is, my hypothesis involves the idea that an intention$_B$ (the *felt* intention to act) is an interpretation of the intentional$_A$ immersion dynamic involved in paring down degrees of freedom for action (i.e. making choices). In other words: the *phenomenology* involved in being immersed in the structure of constraints and enablings realised in the *interaction* (dynamic interplay) of one's own body and the physical and

social world (intentionality$_A$) can be interpreted/explained as having the intention$_B$ to do such and such.

This intention-dynamic involves resonating along with the normative fields already present in the environment, interacting with them and exerting modifying influence. This resonance consists in an interpretational dynamic of one's own mode of embeddedness in this compound normative field: a continuously shifting push-pull balance of embodied emotions (originating in embodied processes, i.e. to be expressed at least partly in terms of M-space), as well as constraints and enablings in the environmental (P-space) and social (S-space) sense. Interpretation here is a form of action; it involves paring down the agent-environment interaction-system's extant degrees of freedom. The result of this interpretational dynamic is the C-space dispositional array. That is, the agent-world resonance involves exactly the same constraints-and-enablings-dynamic that generates, via dynamic dimensioned realisation, an agent's conceptual dispositions.

Hence, my suggestion is that the agent's interaction with such normative structures evokes the development of a behavioural dispositional array of a specific nature: C-space. Having concepts, i.e. being disposed to (re-)act in a certain way in a specific context, is the structured answer to normative evocations (affordances) by the physical and social environment.

To reiterate my remarks from section 6.6, concept attribution occurs when the other agent meets the criteria, i.e. (in terms inspired by Brandom) an agent obtains entitlement to be attributed certain concepts by exemplifying the appropriate kinds of behaviour that can be interpreted as endorsing the possession of these concepts. What is and is not 'appropriate' in this context will, in general but not exclusively, be defined in terms of *expectations*, i.e. whether the other's behaviour fits the profile we feel it should fit. If an expectation is violated, there is a chance some amount of reflection on the part of the observer will unearth a deviant justification of the observed behaviour, but I submit that in most cases the 'this was not what I expected'-reaction will suffice to disallow the wholehearted attribution of some concept.

This reaction, this (embodied) feeling, is the basis of being entitled to being attributed certain concepts, both reflexive (interpreting oneself as acting in a way that validates the ascription of a specific concept) and attributive (having such properties and acting in such a way that others will be disposed towards describing an agent as having a particular concept).

9.3 - Impredicative Loops

I am fully aware of the fact that the way in which the concept 'concept' - and even cognition in general - has been treated throughout this book, smacks of circularity. There is repeated talk of mutual and/or holistic dependence of concepts on other concepts, co-constitution of properties and processes,

causal stories with no clear beginning or end. My claim in this section will be that this is necessarily so.

Hofstadter (1979, 2007), discerns what he calls 'strange loops' in a wide variety of objects, processes and structures, most famously in the art of M.C. Escher, but also in DNA, and even 'the self' as such:

"And yet when I say "strange loop", I have something else in mind - a less concrete, more elusive notion. What I mean by "strange loop" is - here goes a first stab, anyway - not a physical circuit but an abstract loop in which, in the series of stages that constitute the cycling-around, there is a shift from one level of abstraction (or structure) to another, which feels like an upwards movement in a hierarchy, and yet somehow the successive "upward" shifts turn out to give rise to a closed cycle. That is, despite one's sense of departing ever further from one's origin, one winds up, to one's shock, exactly where one had started out. In short, a strange loop is a paradoxical level-crossing feedback loop." (Hofstadter 2007)

Another way he parses this insight is by describing processes in terms of *tangled hierarchies*, where different sub-processes at various levels influence and lock into eachother, in such a way that following the causal chain eventually puts one back at the starting point. (Part of) the problem resides in these processes' self-referential structure, resulting in cases in which it is unclear what causal story needs to be told to explain it.

Chemero and Turvey (2007)[NOTE 82] highlight *Russell's Paradox* as a scenario that contains such logically problematic references. Consider the following: the barber shaves all and only those who do not shave themselves; who shaves the barber? Attempting to answer this question results in a vicious circularity: if the barber shaves himself, he does not shave himself, and if he does not shave himself, he does shave himself. Russell introduced the Vicious Circle Principle to disqualify such sentences from being considered coherent.

In a similar fashion, Poincaré (1906) banished socalled *impredicative* definitions from mathematics. *Predicativity* in mathematics and logic involves disallowing "(...) any set S that contains members m definable only in terms of S, or members m involving or presupposing S." (Chemero and Turvey 2007).

That is, a definition is said to be impredicative if it references itself, e.g. when the properties of a given member of a set depend on the other members of a set. For example: whoever else is in the room does not change the fact that Paul is 1.86m tall, but it might have consequences for whether he is the tallest person in that room or not. When Steve, who is 1.88m, enters, Paul is still 1.86m, but he is no longer tallest. This example (adapted from Chemero and Turvey 2007) demonstrates one of the complicating factors involving impredicative definitions, namely that what they pick out is highly *context-dependent*.

Despite the fact that Poincaré wanted mathematics to have nothing whatsoever to do with impredicativity - and mathematics is the basis of physics, our primary tool with which to describe and define the world -, there are reasons to believe that impredicativity is the *norm*, rather than the exception. Rosen (1991) believes that models of living systems are almost all impredicative. Due to space considerations, I cannot go into the formal reasons of this claim, but one argument Rosen uses is based on Gödel's incompleteness theorem:

"(...) in any consistent system able to produce simple arithmetic there are formulae that cannot be proved within the system but which are seen to be true from outside the system" (Chemero and Turvey 2007)

For Rosen, this suggests that it is impossible to remove all self-referring loops from formal systems, doubly so for the models we utilise to describe living beings. Hofstadter (1979, 2007) too picks out Gödel's Theorem to illustrate his 'strange loops'-idea.

Rosen's thesis about the impredicativity of models of living systems aligns rather neatly with the notion *autopoiesis*, the self-sustaining, self-generating activity of living cells that was conceptualised by Maturana and Varela:

"An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in space in which they (the components) exist by specifying the topological domain of its realization as such a network." (Maturana and Varela, 1972/1980)

Autopoiesis is a central notion in quite a few $E_{(i)}C$ theories. It should come as no surprise that other notions inherent to $E_{(i)}C$ also show signs of impredicativity. After all, if $E_{(i)}C$ is anything, it is an attempt to provide an ontology of cognition as an aspect of the *living organism* as he acts in his environment.

One of those impredicative notions, and an important one at that, is *affordance* (see also sections 4.5 and 8.2). Chemero and Turvey (2007) highlight two ways of explaining the notion 'affordance', one represented by Turvey's (1992) Dispositional View, the other represented by Chemero's (2003) Relational View[NOTE 83].

The dispositional view utilises the affordance/effectivity balance (see also note 26):

"Affordances are dispositional properties of environmental objects. Although they have occurrent causal bases, they are definable only in terms of

complementing effectivities. Effectivities in turn are dispositional properties of animals that have, but are not identical to, causal bases and are definable only in terms of complementing affordances." (Chemero and Turvey 2007)

The Relational View is subtly different:

"(...) although effectivities are typically considered a technical term for abilities, Chemero (...) holds that abilities are not dispositions. Instead, abilities are a variety of functional property of animals, in that they are normative and can be exercised well or poorly. Despite this difference from effectivities, abilities are also defined in terms of affordances. An ability is a relation between an affordance and an activity; having the ability to catch a ball means actually being able to catch balls (an activity) when ballcatching affordances are present. We can understand Chemero's model, then, as saying that token affordances are ordered pairs of particular abilities and particular situations, where those situations are very complex objects that include animals and physical features of the environment. Token abilities, then, are ordered pairs of particular affordances and particular actions." (Chemero and Turvey 2007)

It is very interesting to note that both views of affordances require impredicative definitions: they both feature complex recurring relations between properties of agents and environment, and these properties (as well as the relations) can shift on a moment's notice, with possibilities for action suddenly appearing as the agent moves and interacts with a dynamic environment, and disappearing just as suddenly.

These examples (Russell's Paradox, Gödel's Theorem, autopoiesis, affordance) suggest impredicativity is a highly complex and complicating, but also common feature. My suggestion is that the concept 'concept', as I have developed it throughout this book, is also defined via impredicative structures. The interrelatedness of the spaces of the RM involves complex mutual dependencies and definitions, and this contributes to the fact that the very definition of 'having a concept' contains several impredicative loops.

To see this, consider the following definition, adapted from the one provided in section 6.3, and reiterated in section 9.2 above.

Having a concept means:
being disposed, in a particular context, to act in such a way that concept-ascription by oneself and by others is justified.

Taking this definition apart, we can see the following recurrent structures:
being disposed, in a particular context : this means *embodying* and *being embedded in* a specific affordance-effectivity-dynamic;
concept-ascription by oneself : requires the self-referential activity of *self-interpretation*;

concept-ascription by others : this involves a very slippery dynamic of changes in another's behaviour in response to your own, which yield new social affordances, hence elicit new effectivities (and new dispositions), and influence self-interpretation (hence self-directed concept-ascription);
concept-ascription is justified : this is the case when when concept-ascription yields better explanations of an agent's actions. Here we see multiple loops: the first is that providing explanations requires concepts. The second loop is that these explanations are judged in terms of appropriateness, and this too is a concept-dependent notion: concepts are required to gauge an action's appropriateness.

In addition to these circular structures, the following aspects of concepts also implicate loops: a specific set of dispositions (i.e. a disposition to act in a certain way if just action is required, or the disposition to provide a particular explanation of 'justice' - see section 6.7) depends on the associated concept's enslaver, which depends *diachronically* on earlier encounters with similar affordance-effectivity dynamics, including the proto-conceptual disposition profile belonging to this specific conceptual ability (i.e. the neurophysiological yield and the categorization propensity founded on it), and it depends *holistically* on other concepts (i.e. some concepts can modulate the intensity or application domain of other concepts).

I submit that the RM forms a contribution to illuminating the role of concepts in this impredicative tangle of properties and processes (namely, they serve to enable mental state ascription on the basis of embodied properties, and to *explain* behavioural profiles), and shows how concepts are related to the the agent's body as it is embedded in its environment, i.e. the 'organic substrate' (namely via dynamical dimensioned realisation), as well as how they are related to behaviour (as mentioned, they perform a specific explanatory role).

9.4 - Concept Individuation

RM is mainly a theory about the concept 'concept', but it does provide us with some tools to help us in individuating concepts, i.e. saying what a specific concept means. However, RM's similarity to empiricist theories of concepts (see also section 10.3) might throw a spanner in the works.

After all, RM contains an empiricist streak in the sense that perceptual contingencies (for instance, the neurophysiologically defined response patterns to chromatic stimuli) are claimed to lie at the very foundation of conceptual ability (see chapters 4 and 5). An important line of criticism against empiricist theories of concepts involves the difficulty of getting from that basic perception-based preconceptual structure to the rich technicolour pandaemonium of homo sapiens' conceptual system.

For instance, Laurence and Margolis (1999) criticise noted empiricist John Locke's conviction that basic Ideas can be extrapolated into full-blown

concepts. Locke (1690/1979) attempts to provide a reduction to perceptual primitives of the concept 'lie' like this:

"1.Articulate Sounds. 2. Certain *Ideas* in the Mind of the Speaker. 3. Those words the signs of those Ideas. 4. Those signs put together by affirmation or negation, otherwise that the *Ideas* they stand for, are in the mind of the Speaker. I think I need not go any farther in the Analysis of that complex *Idea,* we call a Lye. What I have said is enough to shew, that it is made up of simple *Idea*s: And it could not but be an offensive tediousness to my Reader, to trouble him with more minute enumeration of every particular simple *Idea,* that goes into this complex one; which, from what has been said, he cannot but be able to make out to himself."

Laurence and Margolis object that this still does not answer the question, even though Locke would hold it to be obvious from what he has already said: the tediousness of providing a more detailed explaination would not only fail to be offensive, but would in fact be quite welcome, not to mention absolutely material to the empiricist's argument.

My suggestion is that the magic ingredient that is missing from this argument is embodied and embedded *behaviour*. To be a little bit more specific: what a concept is, according to RM, is locked in by the conflation of a number of different processes, chief amongst which is a specific behavioural profile, displayed by an embodied agent who is embedded in a particular environment. This behaviour is the agent's attempt to navigate the affordances he finds himself confronted with, given his own embodied abilities and limitations. Let's call this behaviour 'enacting a particular concept'. Someone who chops down a tree does not display behaviour that is congruent with that aspect of the concept 'tree' which describes a tree's role as a shadow-giver, but he might instead be focusing on the possible tree-roles 'fuel' or 'object blocking my path'.

I think this brief description already provides the core of RM's idea with which to answer the question where a concept's meaning (some would say 'content') comes from. That is: Locke's description above is too static, i.e. not nearly dynamic enough. RM, instead, can say that embodiment provides the basic perceptual structuredness that the empiricist refers to. Embeddedness instantiates many interaction opportunities ('enablings') as well as limitations ('constraints') for the embodied system's perception-based dispositions to manifest themselves, and behaviour is the dynamic interplay of these two constellations of properties. This is not a linear exchange of information, but a multi-tracked, interlocking back and forth of influences and processes which collectively realize temporarily stable structures: the dynamic self-organization of behaviour. And of course, one of RM's central claims is that concepts (as belonging to specific behavioural profiles) arise from exactly these kinds of dynamics.

A very basic principle of evolution is invoked when the dynamics arising from the mutual attunement of basic stimulus-response pairs clash with

environmental constraints. When a simple system's habitual response to some stimulus fails to achieve the habitual result, this generates pressure for the system to either adapt or go extinct. The very essence of behaviour is *having options*: exhibiting behaviour is paring down degrees of freedom (e.g. the way in which my environment and myself co-constitute a particular affordance-effectivity structure allows me to go both left and right, but I pick only one of the available options to actually *do*). Evolution is, at least in part, that these degrees of freedom change, creating new forms of self-organization, new resources with which to capitalise upon newly evoked options.

So, RM says that having a specific concept means standing in specific kinds of relations to the environment and all relevant objects and agents in it. That is, an agent, with an array of embodied properties and dispositions is embedded in a particular context, hence is constrained and enabled by properties of that context, in such a way that a specific kind of behaviour is elicited - behaviour which counts, to the agent's conspecifics, as expressing the possession of said concept. This is the general dynamic structure expressed in Dynamical Dimensioned Realisation: the mutual constraints and enablings of various processes at various scales (granularities) conspire to yield a particular kind of behaviour (agent-environment interaction) which can be described in terms of the agent having specific concepts.

Now, there are different ways of going about the project of specifying the meaning of concepts, based on these considerations. For instance, perhaps it is possible to individuate a particular concept that an agent enacts by providing a list of propositional attitudes as they are implied by said agent's concept-informed behaviour - this would be akin to Peacocke's (1992) approach to concept individuation, which is defined in terms of possession conditions. Perhaps one might even get an above-chance performance if the agent is asked whether he agrees with the propositions implied by his behaviour

However, there is an important problem with this approach: the locking-in of the significance of this behaviour (given that particular context) in most cases *underdetermines* the exact meaning of the concept that is being enacted. This, of course, is exactly the idea behind granularity (see section 6.8). One could even claim that there is a two-way underdetermination: (1) two people enacting what they would describe as the same concept might result in two different behavioural profiles (but probably with the same [intended] result), and vice versa: (2) two people exhibiting the same or highly similar behaviour might not, if prompted to explain, justify that behaviour as enacting the same concept.

Here are two examples, one to support either case:

(1) Suppose two people are given the task to identify a sparrow. Both the concepts 'identification' and 'sparrow' can be understood at various

granularities, hence be associated with different kinds of behaviour. My own behaviour would probably consist of pointing when an appropriate (in my non-ornithologically-educated opinion) subject were to flutter by, whereas an expert might offer a detailed description and demonstration of features while picking out the creature in question.

(2) Imagine two people running a marathon. They might run that distance in exactly the same time, following exactly the same route, but invoke different concepts while doing so: one might associate that marathon with 'winning', or 'getting the gold medal', whereas the other might associate running the marathon with 'getting really tired, because I enjoy that feeling'. Hence we have different motivations, different concepts of 'marathon' within the context of those motivations, but (purportedly) exactly the same behaviour. Here we see that this is once again a granularity issue: if we ask these two people why they ran that marathon the way they did, i.e. we try to dig deeper into the kinds of justifications they offer in defense of their behaviour involving this marathon, we might find out about these different motivations, but as long as we refrain from 'digging deeper' along the granularity gradient, we will not detect these differences.

These remarks would suggest that concept individuation being a problem is necessarily the case, because the $E_{(i)}C$-approach to concepts that RM is means that concepts, as are many embodied actions, are *softly assembled* from extant affordance-effectivity dynamics. In other words: most embodied and embedded action is not programmed, but in some sense opportunistic in nature, unfolding along the path of least resistance, given the particular interaction possibilities of the available objects and resources. The dynamic nature of this kind of process means that what this path is, is likely to be highly dependent on many different variables, hence subtly (or sometimes not-so-subtly) different every time.

Still, RM offers a few tools with which to make some progress towards individuating concepts, even in these difficult impredicative circumstances. For instance, a descriptive tool is listing contributing processes, as affordance-effectivity pairs (functional clusters), defined in terms of their contributions to the agent's behaviour (the ontological core - see section 7.8). These descriptions can be formulated in terms of constraints and enablings of certain processes or properties on other processes or properties, collectively sketching a kind of state space within which certain behavioural profiles are available for the agent to enact: see section 8.5.

For instance, the meaning of the concept 'mountain' can, according to RM, be individuated in terms of, amongst other things, body-based features (e.g. my experience of the visual parallax between horizon and the mountain's summit; the strain in my muscles as I overcome gravity to climb the mountain; the sense of movement and the frosty breeze in my face as I ski down a snowy mountain slope) plus other concept-informing experiences (e.g. high-school lectures about the geographical features of the Andes mountain range; hearing someone use mountain-related metaphors ["When

we rob that bank, we'll make a mountain of cash!"]); and/or memories I might have of all these things. These are not so much features of the concept itself as they are features which collectively constrain and/or enable (i.e. realize) a particular concept-expressing behavioural profile - exactly according to the RM model's Dynamical Dimensioned Realization (expressed in figure 24). All these properties and features can be described as depending on other experiences and bits of knowledge that I can extract from this concept's jurisprudence (see section 6.6), each of them constrained in meaning and relevance by other experiences and knowledge. And this concept, in turn, enables and constrains the meaning of other concepts. Suppose that I once barely survived an avalanche. My concept 'mountain' might, because of this experience, be associated with a phobia of heights or unsteady-looking mountain slopes, hence disallow any form of enjoyment occurring on or near mountains.

A much lower-grained version of such an account is invoked in the attribution of concepts as a strategy for explaining behaviour. Recall from sections 8.4 and 9.2 that describing someone else as having a particular set of concepts, and you yourself behaving as being entitled to receive certain concept-ascriptions, is all part of the social game of explaining eachother's behaviour as fitting into a collectively specified normative structure (e.g. a particular ethical system). What is needed for this to occur is to infer, based on observed behaviour, some of the ideas mentioned above (e.g. "he acts scared when on the slope of a mountain, so that must mean that he intensely dislikes being on a mountain, possibly due to some traumatic experience."). Whether or not these inferences are correct down to the very last detail is, in many cases, of lesser importance. What matters is that those inferences provide a sufficient explanation and/or prediction of the other agent's behaviour, analysed at the granularity that is warranted by the situation. Whether I dislike being on a mountain because I once survived being in an avalanche, because I get nauseous due to some deficiency of my equilibrium apparatus or because I simply dislike the smell of fern trees is not that important when the net result is that my friends do not ask or force me to come to the mountains with them. The British say that the proof of the pudding is in the eating; in this case RM can claim that the meaning of the concept is defined in the doing.

9.5 - Epistemological Implications

The ideas about what concepts are expressed in RM, i.e. how we acquire and 'enact' concepts, has certain epistemological implications. In this section, I will say a few things about the nature of knowledge and explanation that flows from the claims of RM.

Marchionni (2008) makes a distinction between two ways in which pairs of explanations - of the same phenomenon or process, but at different levels, say the macro- and micro-level - can complement eachother. If two such explanations are *weakly complementary*, each offers a complete story that contains at least some information that the other does not. For instance: if

certain conditions are met (e.g. mental states are multiply realizable, the hard problem is accepted), a microphysical account of brain activity tells us things that cannot be captured in psychological terms, and vice versa. When two explanations are *strongly complementary*, this means that both stories mutually complement eachother, their integration resulting in a more complete, better explanation.

RM hopes to achieve strong complementarity, as Marchionni envisions it. That is, RM states that contributions from various disciplines and explanations that are focused on various levels (e.g. macro- and micro-levels) of the process in need of explanation should be integrated to yield a better explanation. The explanations from the various approaches, possibly relevant to aspects of the process at various granularity levels, can therefore only be partial explanations: they are contextually appropriate, but here this also means that they are likely to be merely *aspectually* appropriate, i.e. relevant to part of the process, not the entirety.

In RM, the granularity gradient (see section 6.8) defines a hierarchy of descriptions and explanations of levels of constituent processes - e.g. a physical description of a tree (i.e. in M-space) might be given in terms of concepts associated with ecology, biochemistry, microphysics and so on. The word 'hierarchy' is not meant to imply a value judgment pertaining to which level is ontologically most important or most basic (see also section 8.3.2): which explanatory level is relevant is context- and question-dependent. This does not mean that a description of the tree at the biochemical level (e.g. how the leaves' chlorophyl yields energy under the influence of sunlight) is necessarily more *detailed* than a description at the ecological level (e.g. how the tree as an organism functions in its environment) - either account might be very complex and expansive -, but that there is a difference in the scope of the description, and the operational scale of the kinds of entities that are invoked.

Given all these sub-accounts at different granularities, RM offers an explanatory model that involves context-dependent *centripetal* complementarity: the various partial explanations are composed with an explicit focus on the aforementioned ontological core (the embodied agent as he is embedded in his environment). That is, the kind of explanatory pluralism to be found in the RM model involves multiple explanations that cover different aspects of the same complex agent-environment interaction dynamic. In the end, they are about the same 'thing' (namely the agent in his context), but they describe parts, roles or properties of this thing in different ways and possibly at different levels of granularity. The goal is to have all thse different partial explanations complement eachother, and for that to be possible there is likely to be some need for translation, sometimes even for disciplines which pertain to similar phenomena: a psychological account of 'experience' is likely to be different than a neurophysiological account. My suggestion is that we try and redescribe all relevant processes in terms of their contributions to (i.e. the constraints and enablings they generate for) the properties and processes of the embodied, embedded

agent. In other words, the claims of the disciplinary tributaries are to be defined in terms of their contributions to the description of the ontological core.

To reiterate, RM claims that the integrative force of discipline-specific explanations is due to them all being related to an explanatory anchour, in the form of the *agentive* conceptual description. Recall figure 24, which expresses this idea most compactly: the properties and processes involved in the agent-environment interaction dynamic are *collectively* characterised in terms of a concept-based description.

This is why RM would be at odds with the kind of neurophysiological lay-talk that surfaces rather too often in popularising publications (of the kind that is heavily criticised by Bennett and Hacker, 2003): the notion that (a specific region of) 'the brain' perceives, feels, decides, hears and so on. Such descriptions limit the applicability range of these notions (perceiving, feeling, etc.) to such an extent that their intended meaning dissolves. The descriptive and explanatory power of these notions is greatly increased if they are applied to the agentive level rather than the neurophysiological, so the argument to explain at least mental terms at this level can be understood as a form of the principle of charity: they do their best work at that level. The fortifying argument to have the agentive level - the context of the person - be the ontological core, i.e. the primary explanatory nexus, is the contention that this is the Archimedian point on which all knowledge-gathering hinges: *we* attempt to understand - not our brains nor our constituting atoms, but we *as persons*, interacting with other thinking, perceiving and feeling persons. The context of the person is the context of our phenomenology: it is the inescapable viewpoint from which we attempt to understand ourselves and the world.

= =

[SUMMARY of chapter 9 AND PREVIEW]

In chapter 9, I suggested that very basic bodily syntax can help establish practices of participatory sense-making: the embodied interactivity of agents is an example of the social co-construction of meaningful interaction profiles. Recognizing such patterns lies at the root of concept-ascription, our social game of being entitled to being attributed certain concepts, both reflexive (interpreting oneself as acting in a way that validates the ascription of a specific concept) and attributive (having such properties and acting in such a way that others will be disposed towards describing an agent as having a particular concept).

There is a fundamental circularity involved in using concepts, for instance in understanding oneself as using concepts as measured against the socially defined criteria of concept-possession, which we ourselves help establish. In this chapter, I described such circularity in terms of a broader phenomenon, which involves understanding living, autopoietic systems as

implementing so-called impredicative loops. The Radicality Manifold as an $E_{(i)}C$-appropriate theory of concepts fits in with this idea.

RM suggests that it should be possible to individuate concepts in terms of the various constraints and enablings that are posed by the experiences and bits of knowledge contained within a concept's narrative jurisprudence. While explaining the behaviour of others, we can attribute concepts to them which we might individuate (in a practical sense) by making inferences about such experiences and knowledge at a coarse-grained level.

One of the epistemological implications of the RM-model is that the explanatory focus is placed squarely on *the agent as a whole*: the various partial explanations associated with the separate spaces of the RM are composed with an explicit focus on the embodied agent as he is embedded in his environment, so these explanations all cover different aspects of the same complex agent-environment interaction dynamic.

In the tenth and final chapter, the RM-model will be evaluated. There will be an attack on the very basis of the model, phenomenal colour space, and the model will be compared to two similar approaches: Peter Gärdenfors' conceptual spaces and Jesse Prinz' concept empiricism. The RM-model will be applied to a concrete case - concept-based early childhood education - where it will help offer some structuring suggestions. In chapter 10 it will also be determined to what extent the RM-model meets the desiderata on a theory of concepts as formulated at the end of chapter 2.

= =

[10 - Evaluation, Application and Conclusion]

In this chapter, I will evaluate the RM-model. First, I will discuss a line of reasoning that attacks the very basis of the SToCC/RM-framework, phenomenal colour space (section 10.1). I will also include a discussion of Peter Gärdenfors' 'conceptual spaces'-account (section 10.2) and Jesse Prinz' 'proxytype theory' (section 10.3), two theories that are similar in some ways to SToCC/RM, and explain why SToCC/RM is different. Then, section 10.4 contains a description of the application of the RM model to a concrete case - a concept-based curriculum for early childhood education. The sections that wrap up this chapter - and the book as a whole - concern the ways in which RM meets the criteria Prinz' imposes on a theory of concepts (section 10.5), and some concluding remarks (section 10.6).

10.1 - Eliminating Internal Geometric Spaces

First, a potentially serious counter-argument. DeCock (2006) suggests we should adopt an eliminativistic position regarding internal (phenomenal) metric spaces, where the structure of this space is supposed to characterise a perceiver's internal representational content. He perceives the *metric* aspect in particular to be problematic: it is terribly difficult to define the supposedly metric properties and values of such spaces in a rigorous manner, based on introspection of phenomenal experience. An argument in favour of this claim is the fact that there are many incompatible ways of modeling the supposed metric of colour phenomenology: various kinds of spindles, spheres, cones and pyramids, in addition to irregularly shaped forms. Furthermore, this metrical structure is not reflected in the perceiver's neural activity.

Because of these problems, DeCock feels justified in claiming that these internal space constructs might be (or might have been, in a historic sense) heuristically useful, but - like aether and phlogiston - they are ultimately redundant in light of much better theories. That is, physical or psychophysical spaces, defined in terms of objectively measurable variables, can do all the work these internal phenomenal spaces are supposed to do. If any philosopher or psychologist remains steadfast in his desire to use phenomenological spaces, DeCock's advice would be to have these spaces be *topological* rather than metric, and leave the 'metrics' to psychophysical theories (which could, if needed, mould their models in terms of spaces).

In his paper, DeCock affords explicit attention to phenomenal colour space, and to Gärdenfors' 'conceptual spaces'-account. RM implies a view that is related to both of these examples: in chapter 6, I explained the notion of a conceptual space as an extension of phenomenal colour space, and some aspects of the RM-project as a whole align with elements from Gärdenfors (2000)[NOTE 84].

Psychophysical spaces as such (like the CIE chromaticity diagram) - in the case of colour recording the properties and interrelations of the perceived object's surface, of the light and its source, and of the perceiver's retina and visual processing areas in the brain - need not rely on introspection data. However, regarding phenomenal colour space, the orthodoxy amongst supporters appears to be that these externally measurable colour spaces possess the structure they do because they are derived from the structure of internal phenomenal colour space. As noted, DeCock's criticism now focuses on the impossibility of assigning metric properties to this internally 'perceived' structure; for one, it is nigh-impossible to provide an accurate appraisal of 'distances' between colours in this internal space.

Despite the parallels between aspects of these examples and certain claims contained within the RM-model, I believe that RM is not in danger of falling prey to DeCock's arguments. His chief worry is that internal spaces are usually awarded a particular metric structure, which suggests an accuracy that introspection is simply unable to yield. Obviously, the 'represented' properties, relations, values and processes 'encoded' in P-, S-, M- and B-space are all physically/objectively specifiable, either in terms of the properties of the agent's physical or social environment (P- and S-space, respectively), his body (M-space), or the dynamics inherent in his actions (B-space), so there should not be a problem there. Rather, the quarrel would have to involve the status of C-space, the closest thing to an 'internal space' in the RM-model. However, in that case the crux of RM's defence should be that the structure of C-space is inferred in narrative/justificatory terms from the agent's cognitive, locutionary and bodily behaviour. In other words, the relations in an agent's conceptual space can be reconstructed by charting said agent's actions, with as an important subset the kinds of justificatory accounts he provides when pressed to explain his use of a particular concept.

The claim that there is some phenomenal experience involved in the process of generating perceptual judgments (which is a claim that I *do* wish to endorse) need not imply that the structure of phenomenal space is metric in character. That is, I would say the phenomenal aspects of sensory concepts are indeed *functional*, in the sense that they play a role in the distinction tasks that can be used to determine the structure of C-space: the concept 'apple' includes the notions 'ripe apple' and 'unripe apple', and this division is based, at least in part, on the phenomenal 'charge' of the colours associated with ripeness and unripeness. However, I feel the relations between these 'phenomenal charges' are not specifiable in terms of a rigorous metric, mainly because any and all knowledge we have of these relations (i.e. the structure of phenomenal 'space') can only be aquired by way of an interpretation of our own sensorimotor responses - that is, an a posteriori reconstruction of a sensation that simply does not occur in a neatly measureable form, namely awareness of agency. Hence, this interpretatory process will not and can not result in the kind of rigorous internal metric which DeCock qualifies as unattainable.

Surely, the hue-cancellation experiments carried out by Hurvich and Jameson (1957; see note 16 for a description of these experiments), and similar enterprises, can lay bare some of the regularities of colour perception. However, I would tend to agree with DeCock that these experiments say little about the *metrics* of inner experience. Rather, I would wish to hypothesize that they say something about the dynamics of *environmentally embedded behaviour*, i.e. about the kinds of judgments agents of a particular kind make when placed in specific circumstances.

DeCock has fewer qualms with Gärdenfors' 'conceptual spaces' approach - that is, under a non-realistic interpretation of such spaces. In note 51 of his (2006) paper, DeCock states:

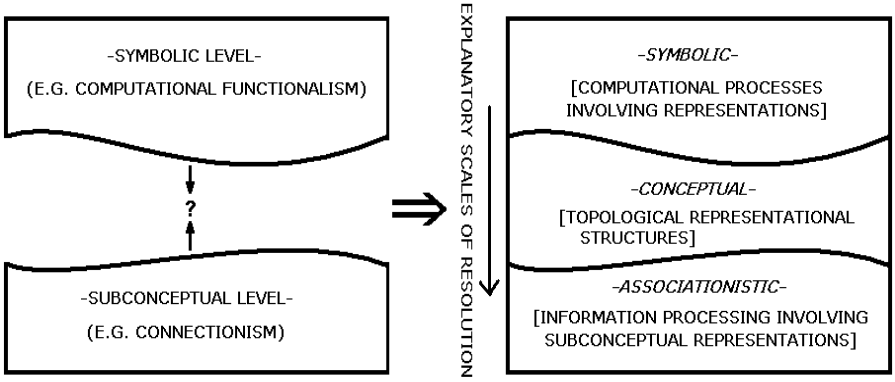
"51. Gärdenfors's central claim is not related to any part of the discussion of this paper; he tries to bridge the gap between symbolic and connectionist approaches by means of geometrical structures. With respect to the philosophical status of his conceptual spaces; he is cautious: "my instrumentalist standing means that I eschew philosophical discussions of how "real" conceptual spaces are. The important thing is that we can *do* things with them." (2000, p. 31) In this paper, it has been argued that realism about 'conceptual' spaces is untenable."

More about how Gärdenfors' theory and SToCC/RM compare can be found below. For now, I can state that the SToCC/RM-approach is not at all incompatible with DeCock's comments, but not for a lack of realism. My claim would be that concepts are real, but not because they are internally tokened in the classical sense (i.e. as an internal, symbolic representation); they are real because concept-involving behaviour is real. And as a result, C-space, on the SToCC/RM-approach, is real: it is a dispositional array. C-space is not exclusively internal, and does not rely solely on representations for its constitution, hence it is as such perfectly specifiable in terms of properties and relations in a way that does not contradict DeCock's argumentation, as explained above.

10.2 - Gärdenfors' 'Conceptual Spaces'

In his 'Conceptual Spaces: The Geometry of Thought' (2000), Peter Gärdenfors develops an intriguing and conceptually fertile contribution to cognitive science. His explicit intent with the 'conceptual spaces'-model (CS) is to bridge the gap between descriptions of cognitive phenomena in terms of, on the one hand, *symbolic* representations, and, on the other hand, *associationistic* representations. The former level of modeling representations involves viewing cognitive systems in terms of computational processes (symbol manipulation), the latter level involves the way in which different informational streams are connected - an example is the subconceptual activity of neural networks propounded by connectionistic theories (i.e. mental content is represented in the activation of a large number of neuronal units).

Gärdenfors suggests his conceptual space should be inserted as an intermediate layer, yielding the following three-tiered model:



[Figure 26: Gärdenfors' 'Conceptual Spaces' account: three levels of representation (in modeling cognitive systems)]

In short, CS is introduced as a 'medium scale' theory to link neurology (involving the activation of large numbers of neuronal units) and psychology (involving language or language-like structures) (Gärdenfors, 2000, p. 50).

Just like SToCC, CS is intended as an extrapolation of accounts involving quality dimensions (e.g. the hue, brightness and saturation of perceptual colour space, but also weight, temperature, height and so on). Defining such a quality along an axis means assigning some ordering relation of the stimuli associated with this quality: '*this* appears brighter than *that*'. A conceptual space (in Gärdenfors' sense) emerges when several such axes are aligned to yield a description of the way in which specific qualities are related for some object (or object-observer-pair, if a quality is perceiver-dependent), and the kinds of combinations of values of these qualities that are possible. A property of an object can now be defined as corresponding to a *region* in this conceptual space: a constellation of linked or related qualities, within a particular bandwidth of values for each of these qualities. So, for example, the property of 'being sea-green' corresponds to a tightly clustered collection of values for the qualities hue, saturation and brightness in perceptual colour space.

As an account of the structure of conceptual space, Gärdenfors suggests that Prototype Theory (see sections 2.3 and 6.11.2) fits the bill, or at least tells an important part of the story with its claim that most concepts have a graded structure: some exemplars are more prototypical of a particular concept than others. The kinds of categorizations predicted by Prototype Theory align with Gärdenfors' characterisation of natural properties as *convex regions*[NOTE 85] in conceptual space. Gärdenfors claims that this, the 'CS' way of characterising properties (quality dimensions define a space in which convex regions, possibly with a prototype structure, represent properties), has a number of important advantages, amongst which is the virtue of making many properties perceptually grounded.

Properties are a special class of concepts, the difference being that a property is based on one domain, and a concept on several domains. Gärdenfors claims this distinction has become muddled due to the advent of accounts of both symbolic (e.g. computational functionalism) and associationistic (connectionism) persuasion, because of their use of (first-order) logic. In natural language, properties and concepts usually correspond to different categories of things, namely adjectives (or verbs, for dynamic properties) and nouns, respectively, but in first-order logic, these are all represented as *predicates*. The CS-account can easily accomodate the difference between the two by taking a very literal approach to the idea mentioned above, i.e. that properties correspond to single domains and concepts to multiple (possibly correlated) domains: they can be represented accordingly in conceptual space.

An example of this multi-domain correspondence for a concept is easy to generate: consider the concept 'goldfish', which is linked to different kinds of properties and specific 'values' thereof (for instance a particular colour, shape and size) and to specific characteristic features (like having fins and gills, or being a water-dweller). Some of these properties and features are linked: possessing gills or having a particular aquadynamic shape have a lot to do with the fact that this animal lives in the water. Taking a cue from Prototype theory, Gärdenfors uses weighted representation to model the extent to which certain properties and features are characteristic for a concept.

This yields the following 'general definition of concept representation' according to CS:

"CRITERION C: A *natural concept* is represented as a set of regions in a number of domains together with an assignment of salience weights to the domains and information about how the regions in different domains are correlated." (Gärdenfors 2000, p. 105)

This is the basis of Gärdenfors' CS-model. Some of it is quite similar to RM (although RM was developed independently, mostly as an extrapolation of claims by Jameson and Maund, and as such founded on the same 'quality space'-accounts that Gärdenfors uses - see chapter 5), but there are also significant differences.

The most obvious, and seemingly inconsequential difference lies in the fact that CS describes a structure of three layers with the conceptual level in between - CS performs a mediating role between the symbolic and associationistic levels, in some sense similar to Churchland's 'vector spaces' (Churchland 1995) - whereas the RM describes a more complex web-like structure.

This difference might appear inconsequential, but it is not: in Gärdenfors' theory, a conceptual space constitutes a (vector-based) suite of representations of properties and categories, and as such one of three

possible resolutions at which the representations that play a role in cognition can be modeled (see figure 26 above). The RM, rather, includes C-space as a partial description of a dispositional field regarding the possible and probable actions of an embodied, embedded (and so on) agent. That is, RM supports a different claim, the deviant nature of which is fleshed out below, when the attitudes of CS and RM to *meaning* (semantics) are discussed.

Furthermore, RM is not about representations as such, but about the agent *as a whole* (although representations can occur). This includes, as a corollary perhaps, the claim that 'cognition' or 'mind' cannot be depicted exclusively in one or more spaces; cognition is an aspect of the agent-world-dynamic in its totality, and that is what the RM describes.

Gärdenfors does note that he intends to link CS to the body:

"Conceptual structures are *embodied* (meaning is not independent of perception or bodily experience)" (Gärdenfors 2000, p. 160)

However, this (plus Gärdenfors' subsequent explanation of this claim: Gärdenfors 2000, p. 161) constitutes a rather limited notion of $E_{(i)}C$, namely $E_{(B)}C$, and little more: CS includes embodiment as perceptual grounding, and as the claim that conceptual structures are somehow linked to bodily experiences and emotions. RM, on the other hand, is, in principle, intended to be applicable to *all* flavours of $E_{(i)}C$.

This subtle difference between RM and CS becomes clearer, and intensifies markedly, when Gärdenfors' ideas about semantics are considered. Gärdenfors makes it explicit that he views the meanings of expressions and locutionary acts emerge from elements of a cognitive structure (this, of course, is his 'conceptual space') to be found *in the heads* of language users, plus sociolinguistic power structures. In his conceptual space, basic lexical expressions in a language are represented semantically as natural concepts, basic adjectives as natural properties, basic verbs as dynamic natural concepts, and basic nouns as multidomain, nondynamic natural concepts.

So, language users are to synchronise their individual conceptual structures (and these imply possibly idiosyncratic meanings, that are quite definitely *in the head*, needing no reference to anything external) to attain optimum communicative efficiency. As such, he argues against Hilary Putnam's (1975) meaning externalism[NOTE 86], claiming that conceptual structures plus the aforementioned sociolinguistic power structures suffice for the existence of meaningful expressions, and no reference to an outside world (beyond that power structure, I would say) is needed.

The claim to be defended on the basis of the RM-account is that if 'meaning' is to be found anywhere, it is *in the system as a whole* - that is, including the body and relevant, choice aspects of the environment. The same goes for

concepts or conceptual structures, which, on RM, are not things to be found anywhere, but rather dispositions towards action, locution and cognition. This latter corollary helps clarify that even despite CS's sociolinguistic aspect, RM simply defends a very different kind of idea of what a 'concept' is (i.e. a disposition-aspect of an agent-world interaction dynamic, rather than a mental entity in the classical sense).

Only part of the explanation for the above-mentioned difference with Gärdenfors' view is that in the RM-approach, whatever processes contribute to mental processing are (or can be) distributed across the system, and whatever meaning is, it involves the entirety of the substrate of the agent's mental processing. To ask for an actual location of a meaning, a mental state or a concept (each of these conceived of as an object of some kind) seems to be the wrong thing to do, simply because it confuses a traditional conception of what 'mind' is with an explanatory approach (namely, RM) that takes as its core objective to get away from exactly that obsolete conception. Still, it should be obvious that an important, and positively *indispensible* part of the processes that account for the existence of meaning occurs in the head: brain activity. The other reading of 'in the head', namely as an aspect of an agent's consciousness, can also apply. The point I wish to make cuts both ways in the sense that these processes 'internal to the head' might be important, but still they fail to tell the whole story, and that it is impossible that they are what they are and do what they do in isolation from that broader context of the system as a whole[NOTE 87].

To drive the point home, consider the tools which CS and RM, respectively, utilise to define semantics. In CS, the cognitive approach to semantics entails mapping linguistic expressions (which might be modulated via sociolinguistic power structures) onto a conceptual structure, which is then applied to the semantic content's target object. Hence, the expression's meaning is constructed internally - that is, prior to application, and independent from any (externally definable) truth conditions. Semantic content comes first, and then syntax is chosen as a shape to pour that content into.

In RM, these two cannot be pulled apart: semantics and syntax form an embodied and socially mediated, dynamic whole. If semantics is to be isolated for descriptive purposes, it can be defined as a dynamic, reciprocal mapping of the RM onto itself. That is, to repeat a remark made earlier (in section 7.6), the relevance of these basic-level (embodied/embedded) processes should be described in *personal* terms, i.e. *the embodied agent as a whole, acting in (and interacting with) a particular physical and social environment*. This reciprocal mapping as a characterisation of meaningful, concept-involving behaviour is another example of an impredicative loop, as described in section 9.3.

In conclusion, I wish to claim that although I respect and admire Gärdenfors' achievement, his CS retains too many cognitivist rudiments to be applicable to theories in the $E_{(i)}C$ realm. It supports only a fairly weak form of $E_{(B)}C$,

and advocates a somewhat unstable compromise of an internally based semantics modulated by sociolinguistic forces. I believe RM offers a more versatile and complete account of concepts and cognition.

10.3 - Prinz' Concept Empiricism

One of the most interesting studies on concepts published in recent years is Jesse Prinz' 'Furnishing the Mind' (2002), briefly referenced in section 2.5, where I introduced his list of desiderata on a theory of concepts. His core concern is to connect the mind - at least inasmuch as it involves concepts - firmly to a bodily basis. As an expansion of this approach - which he calls *concept empiricism* - he presents a theory about how these embodied concepts contribute to cognition, called *proxytype theory*.

Because Prinz' main goal - to explain concepts in embodied terms - aligns (at least in part) with my own, his theory deserves a closer look. In this section, I will first discuss Prinz' theory, with an explicit focus on his use of representations, followed by two sections about the main differences of RM with both main components of Prinz' theory - concept empiricism and proxytype theory.

10.3.1 - Modal Representations

Concept empiricism is, in essence, a reworking of the classical empiricist claim that all ideas in the mind derive from information provided by the senses. Concept empiricism is defined as follows: 'all (human) concepts are copies or combinations of copies of perceptual representations' (Prinz 2002). This is mostly a claim about causality, which Prinz calles the 'Perceptual Priority Hypothesis': mental states are caused by states external to the brain, and the causal chain runs through the senses. Mental states being 'copies or combinations of copies of perceptual representations' adds a claim which Prinz calls the 'Modal-Specificity Hypothesis': 'concepts are couched in representational codes that are specific to our perceptual systems' (Prinz 2002). The upshot of this is that the content of concepts is not merely delivered via the senses, but is also specified in terms of the kinds of information specific to the various modalities.

What is interesting about Prinz' suggestion, at least from an $E_{(i)}C$-perspective, is the fact that it combines a body-based account of concepts with a rather 'classical' use of representational mechanisms. In this sense Prinz offers a neo-empiricist account which aligns with the fairly conservative brand of embodiment and embeddedness as it is also supported by Damasio, in that nowhere the very notion of a 'representation', and its role in concept-involving processes, is called into question. Where Hutto would have us remove as much content as possible from concepts (see section 7.2), in Prinz' story they are filled to the brim with modality-specific content, being internal representations of a rather unapologetically cognitivist kind.

In Prinz' story, these representations are correlated with selective neural responses to stimuli: it is possible to isolate neurons or groups of neurons which become most active when confronted with lines or edges of a particular orientation, specific chromatic properties of objects, specific surface textures or auditory properties, and so on. When a person is confronted with an object, specific distributions of neurons become active whose response preferences collectively align with the set of perceptually accessible properties of the object in question. Such response patterns are hierarchical affairs, with neurons or collections of neurons collating information from more basic levels of processing, eventually resulting in activation patterns which respond to more comprehensive representations of the external object or scene as a whole instead of disparate pockets of neuronal activity caused by specific object features. These representations of complete objects or scenes, not necessarily 'images' but more schematic constellations of information about the properties of the object in question, are then stored in long-term memory. This schematic nature of this complex representation allows the representation's owner to abstract away from the particulars of the object token he perceived, and generalise.

One suggestion mentioned by Prinz would be to identify a concept with the entirety of stored information, particular to the object the concept is of, in long-term memory. Because we, generally, are not conscious of each and every bit of knowledge about a particular object when we can still be said to use its associated concept, Prinz judges this suggestion to be unworkable. Instead, he suggests that a concept is a mental representation that is or can be (temporarily) active in working memory. That is, concepts are what Prinz calls 'proxytypes', which might be simple images or highly complex multimodal representations and anything in between, and which can be used in working memory to represent a particular category of objects in the world. In essence, having a proxytype is generating an internal simulation that is similar to the perception one would have if the actual object were in front of you.

These proxytypes are aggregates of more basic representations, corresponding to the various components and features of the object, based on actually experienced images and sensations stored in long-term memory. As such, proxytypes differ from the symbolic representations of computationalist functionalism: they are explicitly modal and concrete, rather than amodal and abstract.

Prinz' theory is interesting and promising, and similar to RM in several ways. However, there is an important difference: the way in which 'representation' is understood and used. In the next two sections I will examine the two main components of Prinz' account - concept empiricism and proxytype theory - in order to highlight that difference.

10.3.2 - RM and Concept Empiricism

Recall that the main idea behind concept empiricism is the classical empiricist notion that whatever ideas are in the mind, will have gotten there via the senses. Hence, according to Prinz, all concepts, even the abstract ones, are somehow recombinations or reworkings of perceptual representations.

RM aligns with concept empiricism in a specific way. That is, RM describes concepts as emerging out of a developmental interdependence of biomechanical and affordance-based dynamics: properties of the body co-evolving with properties and influences from the environment. In the case of concepts and concept-based cognition, an important portion of those influences is indeed absorbed via the senses and results in mental representations. This does not mean that any concept's origins can easily be traced back to straightforward recombinations of perceptual representations. I do not expect Prinz to disagree: he allows extant concepts as recombinations to be, in a sense, metaphorical. In that light Prinz refers to work done by 'cognitive grammarians' such as Lakoff and Johnson (see section 6.9), who, for instance, explain the emergence of an understanding of causation as abstractions of perceived cause-effect pairs.

However, RM allows another kind of influence of 'the world' on the emergence of an agent's concepts, mainly due to the fact the notion 'representation' as used in RM is different from the kind that Prinz insists on subscribing to. RM offers room for agent-environment interaction on many different timescales - including the evolutionary timescale! - to shape embodied predilections and dispositions, hence conceptual abilities. The proposed mechanism is dynamical dimensioned realisation (see section 7.8), and at least some influences on the dynamics of conceptual or cognitive processes were then indeed first in the senses, but in a very *indirect* fashion.

Prinz's point is that whatever we can think, whatever concept we might have, is highly likely to be an amalgamation of sensations and bits of knowledge we have gathered throughout our life, or variations upon the themes set out by those perceptions. My point is that many influences can still be highly significant, despite not being processed to become 'internal perceptual representations'. Even so, such influences do still count as as contributing to the dynamical dimensioned realisation process, and this is where an important distinction between RM and Prinz' theory lies.

I have an example, describing a rather roundabout process that is not entirely uncontroversial, but which does show how cognition (and, putitatively, concepts) can be influenced by many different properties and processes, in non-obvious ways and over long timescales. I do not suppose Prinz would argue against this indirectness-argument, but it deserves stressing nonetheless.

Here it is: in his (1996), Steven Mithen describes how he believes human cognition evolved. An important step in the evolution of a larger brain, he says, came with the *Australopethicus Afarensis* ('Lucy'), starting roughly 3.5 million years ago. There are indications to support the hypothesis of a climatological change resulting in more arid and open environments in Africa 2.8 million years ago, and changes in the early humans' posture (i.e. starting to walk upright) allowed them to adapt to those changes. The erection of the posture of man to a bipedal form required a larger, substantially more energy-draining brain for muscle control and balance, but resulted in a significant reduction of incident light on the body which enabled longer foraging periods without food or water (hence allowing them to function in that increasingly dry and savannah-like environment). A redistribution of functional space in the brain, reducing processing power needed for the feet (signifying the shift from graspers to weight bearers, which are less demanding in terms of brain processing capacity) might have accomodated the evolution of the hands as specialised appendages (to be utilised for carrying and handling tools, for instance). Other changes of note due to the assumption of the new posture would be an increased frequency of face-to-face encounters, sowing the seeds for an expansion of social and communicatory prowess, and the colonisation of a scavenging niche. That is, the disadvantages from spending time in the sun had been reduced, making it possible for the bipedal proto-human to prey on carcasses at times his rival predators needed to dwell in the shade - a greater amount of meat in the diet allowed the digestive tract to grow smaller, allocating more energy to the upkeep (and evolutionary expansion) of the brain while maintaining a similar base metabolism.

Many of these physiological changes, brought about by environmental factors (dryness and less vegetation), influenced the cognitive capacities of early man: increased use of the hands (subtle and precise manipulation of objects), more frequent face-to-face contact (opening up a world of complex social interaction possibilities), and a general increase in brain size in all likelihood allowed more elaborate cognitive processes to take place, with the appropriate associated conceptual abilities.

For instance, quite a bit of social interaction depends on more or less consciously observed changes in facial expression. Watching a painful facial expression, especially of someone you care about, tends to evoke - in some sense - painful or pain-related sensations in you, and tends to elicit caring behaviour: this is an example of the kind of largely automatic processes which play an important role in how you interact with other people. Children who are denied frequent face-to-face contact at crucial periods of their development (for instance in the case of 'feral children'), or who have deficiencies in the autistic spectrum tend to have severe difficulties in acquiring an effective facility for facial-expression interpretation, and their social skills usually suffer because of it[NOTE 88].

If this is right, it stands to reason that 'theory of mind'-related abilities and concepts might not have become nearly as advanced as they are in modern

humans if such an apparently simple development as starting to walk upright had not occurred, and this might not have occurred if the climatological changes mentioned above had not taken place.

Now, this example is controversial because anthropologists do not agree on the factors that contributed to the increase of brain size in early humans, and I do not intend to either endorse or denounce Mithen's suggestion. However, what this example does show is that conceptual changes (or the possession of certain conceptual abilities *at all*, in this case abilities involving particular social concepts) might have to do with a wide variety of interlocking processes, some of them linked to eachother in decidedly non-obvious ways, or via complex causal chains with many degrees of separation.

Obviously, this example highlights how the environment's influence on evolutionary processes creates the preconditions for particular kinds of concepts appearing, which does not rule out the possibility that any concept that one might possess is, in one way or the other, directly related to some occurrent process. 'Theory of mind'-related capabilities might have convoluted evolutionary origins, but if I have an emotional reaction to some facial expression *right now*, that has everything to do with the fact that I am faced (quite literally) with another person *right now*. Still, in that case being confronted with another person is the trigger or enabler for being in a particular mental state, an enabler furthermore which presents a number of constraints on my own behaviour (e.g. the face is happy, so a sad reaction on my part is less appropriate - unless the happy person is my arch nemesis, of course). However, other highly important sets of constraints are offered by my conceptual and neurophysiologial properties, and the evolutionary history which made those properties what they are - i.e. my C-space and M-space properties. For a complete explanation of my mental state, hence my conceptual behaviour, we cannot limit ourselves to any all-too-straightforward variation upon the explanatory theme 'external event --> perceptual processing --> concept' - *that* is the point of RM.

The above also means that, in some way, the representational nature of concepts as Prinz wishes to defend it is called into question. What brand of representation (if any) is in play when many of a concept's properties are diachronically realized in the way described above (or something in that vein), instead of synchronically? Obviously this question hinges on the difference between the causal origin of a concept and the external entity a concept would be taken to represent - my point is that both these processes might be called representational, but the former in particular is not of the standard 'external event --> perceptual processing --> concept'-kind, for reasons explained above. In my opinion, the Dimensioned Dynamical Realization-account offered in section 7.8 does a better job of characterising the active interplay of actions and impressions within a context of constraints and enablings that occurs in the evolution, over long timescales, of concept-involving agent-environment interaction.

There is another issue here. The argument above is about the indirectness involved in the creation of concepts - concepts might be modal, but not necessarily as a result of straightforward perceptual processing. Presently, I intend to tackle the question of the emergence of abstract, non-modal concepts. Now, to reiterate, Prinz's concept empiricism means that all concepts are specifically linked to information as provided by the perceptual system. Prinz *does* have a story about how certain abstract concepts emerge - part of that story is recounted in section 6.9, for some of the mechanisms he proposes (sign tracking in particular) are mechanisms that RM can also utilise.

The main difference between Prinz' idea about the emergence of abstract concepts and the position to I wish to defend, is that according to RM, certain abstract concepts might be modal by heritage, but need no longer be tied to or couched in terms of the phenomenology associated with a particular modality, or any modality whatsoever. Basic concepts (usually expressed in terms of more or less intuitive behavioural dispositions, only experienced or justifiable after the fact) are modal by definition, because they are an (explanatory) link in an action-reaction chain that is sparked by stimuli of one or more modalities. In that case too it might be unclear from which modality the concept hails, exactly because the behaviour in which the concept is expressed is not a reasoned and planned action. Based on these considerations, I would claim that Prinz' modality claim is only easily supportable for concepts in the 'vulgar centre' bookended by the basic and abstract concept types described above: everyday concepts such as 'redness', or 'chair', or 'horse', that are comparably easy to indicate or perhaps even define - which is why the vast majority of examples in philosophical texts about concepts involve exactly these kinds of concepts.

But of course there are many other concepts. Now, Fodor (1981) famously argued against prototype theory by providing examples of concepts without prototypes (see also section 6.11.2). However, these examples were notions of the kind 'all teal-coloured objects heavier than ten kilos west of the Mississippi' and these might, under some sets of criteria, need to be counted as a concept, but it is somewhat unlikely that anyone has ever had this particular concept (except perhaps Fodor). Of course part of Fodor's point can be that philosophical theories need to be precise enough to disqualify obviously ridiculous implications, but I would wish to suggest that the interpreting party has a responsibility too. As a variant of the principle of charity, I feel philosophical theories should be subject to a 'fair use'-policy: if it takes a really far-fetched example to discount a particular theory, this might indicate a failure of the theory's designer to plug one or more difficult-to-reach holes (and the designer should be held accountable for that), but it might just as likely indicate the weakness of the attack.

However, 'justice', on the other hand, is a concept that many people have, in one way or the other, but this is one of a kind of concepts that many theories of concepts have problems explaining. I provided an account for 'justice' in terms of SToCC/RM in section 6.6. Based on that discussion it

becomes clear that the main difference of RM with Prinz' theory, and probably the feature that allows it to provide a better account of such abstracts concepts, is that RM does not understand a concept exclusively as an internal representation. Rather, RM describes concepts as descriptions of structural properties of an embodied and embedded agent's behavioural dispositions. According to RM, having the concept 'justice' does not mean having a perceptually based proxytype-representation of 'justice' - I find it difficult to grasp what this even means. Rather, it means being able to act in a just manner, or being able to recognize just behaviour in others, or being able to provide arguments about the concept, all in contextually appropriate ways: RM defines concepts behaviourally (with a possible representational correlate), rather than purely or mostly in cognitivistic representational terms. These arguments about the concept can be couched in exemplar terms: the enslaver of the concept, which functions as a mnemonic anchour, with its associated narrative jurisprudence. Now, these enslavers are somewhat similar to Prinz' proxytypes, but RM ultimately defines these tendencies to use specific examples in a way that is quite different from the representation-based definition of Prinz' proxytype. That is, RM says that attributing concepts is to provide descriptions of the structural regularities that emerge in detector-mediated agent-world interaction. Or, put another way, concept-use is what happens when sufficiently sophisticated constellations of detectors (sensory organs) are linked in the appropriate ways with effector systems (e.g. limbs) - where this sensor-effector-linkage does not necessarily require a cognitivistic representational intermediary, but rather functions using the kind of $E_{(i)}C$-appropriate representation descibed in section 7.8 (i.e. DDR).

RM's focus on the behavioural aspect in defining concepts rather than the 'representationally convenient property'-aspect selected by Prinz explains RM's smoother fit to abstract concepts, or concepts that somehow require a relation- or role-based explanation. RM is not committed to properties as comprising sets of individuating features of concepts, but differential behavioural profiles. It is perfectly acceptable to count amongst those behavioural profiles the acts of providing role- or relation-based explanations, and/or performing acts that express an understanding of such roles or relations, or even to provide explanations in terms of examples derived from a concept's enslavers. But usually, you express some level of understanding of what the concept 'game' is by joining in and actually playing a game the way it is supposed to be played.

10.3.3 - RM and Proxytypes

Prinz' proxytype theory suggests that using a particular concept involves re-presentation, a kind of simulation, of (parts of) the object that concept is about, and this proxytype representation is composed of (variations on) formerly gathered perceptual representations of the object in question (or of something sufficiently similar). Hence, when using the concept 'dog', the associated proxytype can contain experience-based images, sounds, smells and the like of dogs or parts of dogs: bodily features such as the dog

having claws, teeth and fur, behavioural features such as the dog being prone to running, licking hands, or rolling over, or other features such as the dog barking, and so on. Depending on the way in which the 'dog'-concept is utilised, different amalgamations of 'dog-features' retrieved from memory might come to form a proxytype as it is 'active' at a particular moment. As such, Prinz' theory of concepts is a profoundly representational theory, in a fairly straightforward cognitivistic sense: a proxytype is a mental object, active in consciousness at a particular time. That is, proxytypes are representations that contain subconcepts of properties, and the real properties in the world these are subconcepts of are used to track natural kinds in the real world, and that tracking relation is the anchour to which the internal-to-external correspondence is moored. According to Prinz, a proxytype (or a concept as such) is a reliable category-detector, its features corresponding to properties of real-world objects, hence serving to pick out those objects as belonging to a certain class or category.

My alternate suggestion is that in the vast majority of cases of concept-involving behaviour, we have no images or other imagined perceptual representations of the concepts we are using. When I see a dog bearing its teeth and running towards me, I run away while seeing and/or hearing *that very real dog*. I do not really see what purpose the intervention of a conglomerate of mental perceptual representations - concepts-as-images of 'dog', 'danger' and 'evasive action', perhaps - would have to serve. If my instincts do the job they are supposed to, I will have jumped up and started running before having had the chance to form a fully realised image-like mental representation of the dog that is chasing me. At that point, most representations that I do have will probably be of fences in the distance that I have to jump over - forcing me to plan my pacing for a good leap - and similar objects.

That is not to say that it is impossible that in some cases, you actually do imagine a dog as Prinz suggests his proxytype works, i.e. with an image-like awareness of the dog's appearance, a sound-like awareness of the dog's bark and so on, in the absence of an actual dog in your proximity. Indeed, those imagined bits of perceptual mental content are perceptually based, and they do belong to the concept they are evoked in reference to. In such a case, those 'images' might be used as mnemonic support when asked to explain what a particular concept means to you - this would be a representation-hungry problem of the kind Clark discusses (see section 7.4). However, my claim would be that those instances in which there would be talk of a fully realised 'proxytype' are the exceptions, rather than cases which specify the norm.

Now, what to think of the similarity of Prinz' proxytype to RM's 'enslaver'? Each of these two functions as a stand-in, during a particular instance of concept-use, for a much more elaborate cache of standing knowledge that need not (and in all likelihood cannot possibly) be 'active' every time that specific concept is used.

The main difference is that an enslaver is not (necessarily) a representation that guides or structures occurrent processes, but rather a collection of features that contributed formative influences ('programmed' expresses what I mean, although the computationalistic connotation is awkward) to an agent's behavioural dynamics, and a description or recollection of which can be reconstructed (usually only in part) to be used in a justification of behaviour ('concept use') during or after the execution of the act in question.

In RM, a concept can indeed be described by referring to a list of properties, in justifying one's concept-based behaviour when pressed to do so, and perhaps (some of) these properties can even be represented. However, the main individuating aspect of a concept is the associated behaviour, which is judged in terms of appropriateness relative to the demands of the situation in which the concept is wielded, as well as the aptitude in defending this behavioural choice, either in word or in act.

Hence, in RM a concept is not a thing inside the head or a representation in the mind[NOTE 89], but a way of describing certain structural features of behaviour (including cognition and locution), either that of oneself or of another. That is, having a concept means being disposed to act in such a way that concept-attribution as an explanatory strategy (again, by oneself about oneself, by oneself about another or vice versa) is justified (i.e. produces an acceptable explanation).

So in terms of structure, proxytypes are similar to enslavers, but in terms of what they actually are (i.e. what the concept 'concept' denotes), there are significant differences. To the heart of this difference goes the claim that RM has a different story to tell about what a representation is, and what it does, than Prinz - see section 7.8 for RM's account of representation as dynamical dimensioned realisation.

In contrast with Prinz' account, RM does not suggest there to be an actual one-on-one reference relation between an internal representation (the concept, or proxytype, or whatever) and the external object, but an interaction dynamic between agent and environment, which can be described or explained in part by referring to concept-possession. It should be obvious that Prinz defends a kind of mental realism; the above means RM can maintain a realism of concepts as behavioural dispositions (including the possibility of cognitive behaviour) and explanatory tools, and that might be defined as mental realism - if one would wish to do so - under the $E_{(i)}C$-redefinition of what the notion 'mental state' is supposed to denote (namely, an agentive, multimodal disposition). That does *not*, however, mean that RM is committed to the idea of real internal representations-as-images that refer to extramental entities.

10.3.4 - Dealing with Problems

This difference in the use and definition of 'representation' allows RM to deal with problems that Prinz, according to Markman and Stillwell (2004), has difficulties solving.

For instance, Markman and Stillwell state that Prinz aligns with most theories about concepts in defending a property-based view: concepts are defined in terms of their properties. The problem there, they say, is that many concepts as we actually use them are understood in terms of their roles (such as 'game' or 'job': playing a game or having a job means playing a particular role in a specific context) or are relational ('sister', 'uncle'). The problem is that Prinz' perception-based concept theory is not particularly appropriate in defining these aberrant concept types, because these concept types are not characterised in terms of perception-based, internally represented properties.

RM is not harmed by this objection, because it is not representational in the cognitivistic sense. In RM, behavioural profiles associated with concepts are individuated in terms of their roles in the agent-environment interaction dynamic: exactly the kind of meaningful, situated role-based coherence Markman and Stilwell accuse Prinz of not being able to provide.

Another remark made by Markman and Stillwell concerns the fact that Prinz defends a strong realism about intentionality. They suggest that it would be much more convenient to replace Prinz' realist position about concepts' reference to external entities with a *coherence*-based view: in such a case the defining feature is not the correctness of the relation between the conceptual representation and the external object, but the consistency of one conceptual representation with other representations.

RM offers a suggestion that is somewhat different from the one made by Markman and Stillwell. The main difference concerns - again - the non-orthodox use of representation: RM does not require a representation as an internal stand-in to establish the intentionality of concepts. Rather, RM speaks of appropriateness-of-use of concepts - not truth -, and defines that appropriateness in terms of justification that a concept-user is required to provide, in one way or another, within the social context of using concepts. As such, there is, at least in part, a collective social construction of the reference of concepts, constrained and enabled by non-socially established facts of the matter (physical environmental, and biomechanical/body-based properties) about the world, and the persons that dwell in it. This is exactly the interplay of properties that can be expressed in S-space, P-space and M-space, and the effect that interactional dynamic has, in terms of dynamical dimensioned realization, on the properties of C-space, as explained in section 7.8 and chapter 8. This mechanism creates the meaning-providing coherence desired by Markman and Stillwell, without requiring the kind of representations Prinz uses. In RM, a strong realist

position about (representational) concept reference is traded in for a dynamic, social practice-based pragmatism about the meaning of concepts.

10.3.5 - RM vs. Prinz: Conclusions

In conclusion: Prinz' project is intriguingly daring, and in my opinion very correct in its wish to find an embodied foundation for concept use and possession. At first glance, it might appear to be the case that there is a match between some of his ideas and some of the components of RM[NOTE 90]. I have shown that that match is indeed appearance, for RM's underlying theory - the notion of representation and the very idea of what a concept *is* in particular - is quite different from Prinz' bold account. My criticism of Prinz' approach is rather mild, because I like his theory very much. The most important difference between his idea and mine is the way that 'representation' is understood, with his account being entirely too cognitivist, and my account offering a redefinition that is $E_{(i)}C$-appropriate.

10.4 - Applying RM: Concept-based Early Childhood Education

In this section, I will highlight a concrete application of RM: the ideas encapsulated in RM about what a concept is can fortify or modify some recommendations made by Birbili (2007) about a concept-based curriculum for young schoolchildren. That is, RM's contribution here is an explanation of why certain aspects of such a concept-based educational program need to be the way they are: the properties of concepts themselves suggest the appropriateness of particular educational strategies. These are the main conclusions to be reached in this section: the necessarily embedded nature of concept-wielding agents implies a connectedness of their concepts to many other concepts (the 'embedded manifolds'-idea - see section 6.6); the embodied nature of concept-wielding agents suggests the success of a multi-modal approach to the acquisition of knowledge and abilities (which is most compactly expressed in figure 24; see also section 9.1); and the social dimension of concept use necessitates a critical attitude focused on justification of choices (see section 9.2)

In her (2007), Maria Birbili advocates a *concept-based* rather than *fact-based* curriculum for early childhood education. The epistemological upshot of this approach is that truly practically applicable knowledge is not characterised by or composed of constants ('facts'). Rather, it is much more useful to train children in the ability to see patterns and connections, compounded by the aptitude to utilise those insights to adapt to constantly changing viewpoints and factual claims - a concept-based approach.

One of the problems with fact-based education, says Birbili, is that there is a danger of presenting children with a collection of disjointed and abstract bits of information. Focusing on facts is an approach to information, furthermore, which suggests a certainty and solidity of knowledge claims which is increasingly difficult to support, given the speed with which ideas and theories shift in modern society. Also, a fact-based curriculum mostly trains

children's mnemonic abilities - important to be sure, but limited in scope and applicability compared to other useful abilities such as pattern-recognition and problem-solving.

To avoid these problems associated with fact-based education - fragmentation and petrification of knowledge, and an excessive focus on the lowest level of cognitive ability, namely memorisation -, Birbili supports the development of concept-based curricula. The core of this approach is to shift educational attention from offering information about specific topics to helping children understand the ideas and generalisations behind those topics.

A telling difference betwen the two approaches is this: in the fact-based approach, a greater depth of instruction means providing more detail, hence teaching more facts pertaining to a particular topic; in the concept-based approach, a greater depth of instruction means stimulating a new level of understanding, which includes being able to see interdisciplinary links, similarities and generalisations, and having the capacity to create new insights based on that knowledge.

For example (following Birbili 2007), when teaching young children about the weather, fact-based education might focus on the different kinds of weather there are and the associated descriptors (sunny, cloudy, rainy, cold, wet, dry, and so on), and that people wear different kinds of clothes in these different conditions. A concept-based approach, rather, might see acquiring such factual information not as an end, but as a jumping-off point to have these children consider concepts such as 'change' (the weather can go from sunny to cloudy), cause and effect (people wear winter coats because it is cold, not the other way around), as well as deeper connections such as 'changes can be observed and recorded' or '(seasonal) changes affect people's activities'.

Birbili stresses that using a concept-based educational strategy does not mean that facts are completely unimportant: rather, she notes that concepts emerge from the classification of factual knowledge, and that teaching facts provides 'supporting detail'. I agree that fact-based educational strategies should continue to be used, as factual knowledge forms an indispensible foundation of many cognitive abilities: all cognition needs Archimedean points. However, I wish to make an additional, somewhat stronger claim: teaching even (moderately) outdated knowledge can be useful. This might appear to be in conflict with a key argument in favour of concept-based education, namely the fluidity of information: what is accepted scientific fact today, might be refuted and outdated tomorrow. However, despite potential interpretational difficulties between paradigms (if one follows Kuhn, 1950), I would wish to claim that even soon-to-be-outdated or already obsolete facts can help sketch a background against which to interpret new information[NOTE 91]. However, these fact-based strategies can be effective only when combined with the flexibility fostered by a concept-based program, in which the pattern-recognition-abilities to see connections

between facts and to apply factual knowledge to novel situations are trained.

Recall that the RM-model highlights the interconnectivity of concepts, and that it claims that having a concept entails knowing (having, in some sense, knowledge of) what kind of behaviour, in a particular context, is required to achieve a particular goal. In that sense this knowledge might either consist of consciously accessible insight or exist in terms of behavioural dispositions. Often, these two are conflated in some way: the appropriate kinds of behaviour are often shaped by the knowledge (e.g. extracted from past experiences in similar situations) an agent has of the objects and/or context in question.

This means that acquiring some concept means acquiring the ability to put the knowledge one has to good use in a particular context, even if some aspects of the situation in which the concept is applied are novel. Therefore, concept-based education can and should train children in applying concepts in a *contexually appropriate* manner. This includes the ability to use a concept at the appropriate granularity (e.g. when picking objects to sit on, categorising solely on a more coarse-grained level ['furniture'] will result in impractical selections - a table is in the same general category as a chair, but that coarse-grained similarity judgment will not help us pick a comfortable place to sit), but also at least some acuity at moving 'up' and 'down' the granularity gradient. This will help grasp the ways in which the concept is embedded at different taxonomic scales, i.e. what the concept 'means' due to its interconnectedness with other concepts: the meaning of the concept 'store' can be understood in virtue of an understanding the connected concepts 'buying', 'selling', 'customer' and so on. At first glance, increasing granularity towards, say, the level of physics will not help in understanding this concept any better, but I would claim that, with the advent of Internet-based stores, understanding the contrast with socalled 'brick and mortar stores' benefits from a grasp of the concept 'physical' as opposed to 'virtual'.

In this sense, it is important to make sure that any educational program latches on to the appropriate granularity of the concepts already present in the child's 'conceptual vocabulary'. Birbili states that young children tend to have difficulties understanding the place of a given object's position within a taxonomic hierarchy, so she feels it is probably best to start teaching at the basic level (everyday objects, conceived of in fairly broad categories: chair, table, dog, cat, tree, and so on), and move upwards (furniture, animals, plants) and downwards (folding chair, lounge chair, stool, etcetera) from there. However, here some of Mandler's (2007) findings are important: early on in their category-development, children might have ideas of how the world should be partitioned that differ from what a scientifically appropriate categorization would suggest. For instance, Mandler reports that certain experimental findings indicate that children under 11 months of age do not differentiate between tables, chairs, beds or even kitchen utensils, grouping all of that together in a 'furniture'-like class. My claim here is that connecting

to the child's conceptual structure can help speed up the child's education, but educators need to be certain beforehand what that structure is, because it might not be what they expect.

If the structure inherent to the Radicality Manifold (as a model that integrates constraints and enablings from many different domains, to express the holistic dependence of the meaning of concepts) is in any way correct, it stands to reason that the best way to teach children the concepts and concept-related abilities is to pick teaching tools (both the methods and the materials used in class) that are child-activating and multi-modal: children should be allowed to touch, taste, feel and listen in an active interaction with the objects and processes they are learning about, *in addition to* being presented with more reasoning-and-fact-based ways of approaching the topics. In this way, both the dynamical dimensioned realization base (physical, social and body-based properties) as well as the conceptual dynamic that emerges from that base are stimulated in a structured fashion in the child's educational program. Obviously, this multimodal presentation should be controlled and composed in accordance with the processing ability of the child to avoid impression-overload: in a play-like context, most children will be able to manage quite substantial information streams, as long as those streams are of the right kind (see below for some caveats).

The suggestion is that if children are allowed to find out for themselves what the information in their (text-)books means, as such activating their bodies in experiencing the object of study in different modalities, this will result in well-rounded, multi-faceted concepts: they will already have had multimodal, *experiential* access to concepts that a fact-based curriculum would have only given them abstract, *descriptive* access to.

Even much more abstract matters such as learning how to spell can benefit from such an *embodied* approach. Bosman and Schraven (2008) and Bosman (2008), for instance, report that, with the proper methodology, even supposedly dyslexic children can learn how to spell just as effectively and quickly as 'normal' children. It might not, at this point, come as a surprise that this method, developed by Schraven and dubbed 'ZLKLS' (**Z**o **L**eer je **K**inderen **L**ezen en **S**pellen - in English: this is how you teach children how to read and spell), is profoundly sensitive to the constraints and enablings of the child's ways of being embodied and embedded.

The ZLKLS-program is geared towards preventing the child from making mistakes, as these erroneous spelling methods can be much more difficult to unlearn than to acquire. The method contains four basic components:

(1) - ***multi-sensorial basis***: learning the letter 'o', for instance, is not just about saying 'o' and drawing little circles. Each letter is acted out, linked to a sound-gesture that is in some way similar to the shape of the letter - for 'o', this is making a loop with thumb and index finger, and moving it away from the eye (eye in Dutch is 'oog', which is a word with a pronounced 'o'-sound)

while making an exaggerated 'o'-sound. This might sound silly, and perhaps it even looks silly, but together with actually writing the letter often and from a very early stage in the educational process, these excercises result in a deep absorption of multimodal embodied 'knowledge' of and experience with that particular letter, making the somewhat abstract task of spelling a word much easier.

(2) - ***direct group instruction***: the teacher functions as the example, acting out a required task. This is a way of capitalising upon the resonance effects inherent to bodily syntax, as discussed in section 9.1. A profoundly important aspect of the environment of a child consists of his teacher, and the other children in his class. When these people are collectively involved in executing a particular behavioural pattern, the biological imperative towards socio-behavioural resonance means that each individual's actions tend to gravitate towards the dominant pattern in that group: the example set by the teacher.

(3) - ***orientation basis***: there is a common, clearly delineated goal for the day. At all points throughout the learning process, children are aware of the goal that they are working towards - for instance: learning all about the letter 'p' -, and how particular input (an assignment) fits into that process. This clarity is essential, because it is so often lacking. Case in point: an important constraining/enabling aspect of the way in which the child is embedded in his environment involves information availability. Sadly, in many classrooms this aspect is such that it impedes the development of the more easily distracted students, such rooms often being littered with spelling charts, posters, crafts areas and all manner of other distracting things to look at. Of course, making learning 'fun' is important, especially for children, and hanging posters on the wall can help do that, but 'fun' should not equal 'sensory overload' (at least for some of the purportedly 'weaker' students).

(4) - ***repetition and examination***: repetition of practice assignments every day (including daily examinations as an additional practice moment). As anyone who has tried to learn how to play an instrument (say, a guitar) can tell you: the mantra is practice, practice, and then practice some more, until at some point that formerly impossible to master arpeggio has become 'embodied': it now feels natural to execute it, and you no longer need to make conscious decisions about where to place your fingers on the fretboard. It is not that different for more abstract tasks such as spelling: practice it via the 'embodied method' described above enough times, and after a while all these mnemonic tools are no longer necessary, the correct answer becoming available almost automatically as soon as the problem presents itself. This is not to say you do not use these behavioural tools any longer, but they are now subdued or at least non-conscious.

The ZLKLS-method has had some remarkable success, allowing seven-year-old children that were placed in special needs schools to learn reading and spelling at a speed comparable to that of 'normal' children, clearly

outperforming comparable children in special needs schools that did not use the method - and all this in the same amount of time, i.e. without the need for additional teaching programs (Bosman and Schraven 2008).

When Birbili (2007) states that it is important to offer children many different kinds of experiences, i.e. to have them learn in a multi-modal or multi-sensorial fashion, I would tend to agree, but with the important caveat (based on the remarks above) that these varied impressions should be rationed, and (obviously) attuned to the developmental level of the children as well as the educational task at hand, to help retain the children's focus during assignments.

Adapting this caveat, as well as the other components of the ZLKLS-method, to a concept-based teaching program, one of the main goals of education becomes teaching children to understand the mutual connectedness of concepts, but in a rationed and embodied fashion. RM suggests that the important role of concepts as epistemic anchours (an enslaver as a compact placeholder for more elaborate suite of knowledge and abilities associated with the concept) should inherently be able to accomodate a teaching program in which the exploration of connections between ideas, of the desire to look for and understand the coherence of concepts in their broader contexts, is stimulated. The aforementioned definitional interdependence of concepts at various taxonomic scales is reflected in the fact that RM models the justification of concept use in terms of embedded manifolds. The idea of an enslaver as an epistemic anchour at the core of an embedded manifold suggests that it should also be possible to unpack this compact core. That is, it should be possible to construct, on the basis of that enslaver - that general idea of what a particular concept means -, contextually appropriate assertions that might serve as justification/explanation of one's use of that concept.

What is needed to generate the claims that are implied by the 'enslaver'-shorthand successfully is a sense of *awareness*: a fair, usable evaluation of one's own position within the grid of occurrent constraints and enablings. That is, the correct use and/or justification of use of a particular concept requires careful observation and experimentation. Given the wide variety of contexts and modalities in which many concepts might express themselves, a student's tendency to *tinker with parameters* is very useful: the drive to ask questions about the properties of concepts in situations that are unlike the context in which the concept was learned. This tinkering might, in many cases, take the form of thought experiments: simply wondering 'what would happen if...?', and as such a foundation of factual knowledge wil most definitely be helpful in constraining the child's imagination, but I would suggest that the best way to learn the basics of this tendency towards experimentation is - again - a curriculum in which children can have the hands-on, multimodal, embodied experience described above, but rationed and implemented in a way that is sensitive to the constraints and enablings inherent to the embodiment and embeddedness of these children.

More in general, I believe that a concept-based curriculum will stimulate the child's critical, analytic abilities, stimulating it to assess the validity of arguments against a backdrop of practical insight acquired via first-hand experience. An important byproduct might be that this approach could also foster, in the child, notions of knowledge and truth as dependent on context and shaped by argumentation rather than derived from dogma or extant social power structures. That is, the child will acquire an analytic outlook on life and knowledge, with sufficient hands-on experience to implement a constructively critical curiosity when confronted by other people's ideas and assertions.

I would (once more - see sections 9.1 and 9.2) like to stress the importance of the social dimension, which returns in the description above ("(...) knowledge (...) derived from (...) extant social power structures.").

If you recall figure 24 from section 8.4, focusing on conceptual ability allows for a more natural explanation of behaviour based on a contextually appropriate (embedded) ascription of reasons for action. A consideration of a behavioural profile's embeddedness in its diachronic and synchronic context - the conceptual approach - is needed for a deeper explanation and socially appropriate interpretation of that behaviour. This social appropriateness is key. After all, RM claims that concepts cannot be characterised just in terms of a classification of the world (as in prototype theory) or definitions (in various guises both in classical theory and theory theory), but in terms of *dynamic, contextually dependent arguments*. In this way, having concepts (or more in general: knowledge) becomes a *social* matter, because such arguments need to be accepted, in some sense, by one's conspecifics in a dynamic of providing and assessing justification, as such attributing and being attributed the possession of concepts.

This was the lesson to take away from the discussion of the ideas of Brandom (see section 9.2): having concepts, i.e. being disposed to (re-)act in a certain way in a specific context, is the structured answer to normative evocations by the physical and social environment. This means that part of instructing children in a concept-based fashion means training children in the practices of attributing commitments (being committed to playing the social game, with all it entails), acknowledging endorsements (accepting the behaviour of others as expressing a particular understanding of the world) and undertaking entitlements (underscoring one's own actions as being correct).

Hence, in essence, one of the main educational goals to be reached for children is the evolving realisation of what it means to have concepts, i.e. *what it is like* to be a functioning, constructively contributing part of an evolving, dynamical physical and social interactional structure. The use of the phrase 'what it is like' is not accidental here: I intend to use 'knowing what it is like' in the *cognitive* sense, i.e. being able to produce arguments in support of why such and such is or should be the case, but also - and not unimportantly - in the *phenomenal* sense, i.e. of having actual embodied,

lived-through experiences of being a part of that social and physical interactional structure.

A good way to encapsulate such lessons is to encouch them in *narrative* structures (see section 6.6). This is not only because such structures most closely resemble the way in which children (and, obviously, people of all other ages) acquire experiences (usually we have one experience [or batch of experiences and attitudes] after the other, organised in a somewhat systematic fashion as one situation leads to the next), but also because such structures most closely align with the ways in which we are usually asked to account for our concept use: by telling a justification-providing story of how certain circumstances conspired to necessitate certain actions. Luckily, we have been teaching our children useful knowledge and abilities by telling them stories for thousands of years, and those children have in turn been practicing those abilities by acting out stories for as long as humanity has been around, so in that sense with concept-based education it would be business as usual.

To recap, RM can strengthen suggestions, such as the one by Birbili (2007), to implement a conceptual approach in early childhood education, by appealing to the very properties of concepts (as understood in an $E_{(i)}C$-appropriate way). The recommendations that can be made, based on RM, involve the following claims:

-fact-based education provides children with a comparatively deficient foundation in life, focusing on a low-level cognitive ability (memorisation and recall);

-concept-based education, on the other hand, trains the contextually appropriate apprehension of objects and their situatedness, as well as the child's own situatedness (embeddedness);

-an epistemic implication of the concept-based approach is that many definitions and explanations tend to be interdependent; a multimodal presentation of learning materials highlights such (possibly transdisciplinary) connections;

-RM, then, contributes ideas about how how concepts/enslavers function as epistemic anchours in the extrapolation of knowledge (see section 6.7), the direction of such extrapolation guided by those conceptual interdependence linkages;

-having concepts, or having the ability to attribute concepts to others, is a *social* property, hence concept-based education should pay special attention to the child's ability to *justify* his concept use; as such, concept-based education fosters a critical, analytic attitude.

10.5 - The Final Evaluation: RM and Prinz' Desiderata

I suggest that the case described above, about a concept-based curriculum in early childhood education, shows that RM can be useful as an explanatory tool. In this section, I will support that idea in a more general sense: I will revisit the list of concept desiderata specified by Prinz (2002) and originally introduced in section 2.5, as a final test of RM as a theory of concepts. Recall that Prinz claims (a claim to which I agreed) that a theory of concepts should meet certain criteria involving scope, intentional content, cognitive content, acquisition, categorization, compositionality and publicity.

The criteria involving ***scope*** and ***compositionality*** are linked, in RM. As a descriptive tool to account for concept compositionality, RM suggests the conceptual spaces account (see chapter 6 and on): the interrelatedness of concepts and the emergence of new concept-compositions can be depicted in terms of embedded manifolds that are characterised by an enslavement structure, linked to social, biomechanical and physical properties in dynamically structured ways. Combining and adapting concepts in ways that are most appropriate to the increasing complexity of the evolving agent's interaction with his environment will result in a spectrum of concepts ranging from modal simplicity to amodal complexity; sections 6.9 and 6.10 go into a bit more detail about the mechanisms and strategies involved in this concept acquisition process. To reiterate: the foundation of concept formation is formed by the sensorimotor apprehension of motion; the expansion and refinement of the catalogue of concepts occurs through the correlation of sensorimotor knowledge and linguistic encoding, followed by embodied and embedded (body- and motion-based) crossmodal mapping (analogies), and finally the linkage of embodiment and abstraction (extrapolated interpretation and implementation of phenomenal awareness, and the exploitation of signs)

These ideas also provides RM with the initial explanatory tools to account for concept ***acquisition***. An additional tool is this: the metaphysical structure underlying the connection between agent and environment, i.e. the structure within which such acquisition processes can take place, can be specified in terms of Dynamical Dimensioned Realization, in which social, biomechanical and physical properties collectively and dynamically specify an interaction process of constraints and enablings, giving rise to concept-evolving behaviour (see section 7.8).

Given this representational structure ('representational' in its alternative definition, provided in section 7.8), ***categorization*** as a concept feature can be explained to depend on the constraints and enablings inherent to said structure. Several important examples of such constraints and enablings - the initial impulses to establish a category structure - derive from evolved perceptual response patterns, such as the structure of an embodied and embedded agent's categorization propensities that can be expressed in phenomenal colour space (see chapter 4).

Several components of RM provide details about how the ***publicity*** of concepts can be established: the *granularity* of concepts (see section 6.8) explains how people can come to believe they are using the same concepts, despite differences in use at more detailed levels; the practice of *justification* (see section 6.6) establishes a mutual attunement mechanism, in which discussion partners are required to account for their concept use and reach some sort of agreement or compromise about the meaning of concepts; justification helps establish a *narrative* 'jurisprudence' of concept use, a socio-cultural practice within which the appropriateness of concept use defines an important part of concept meaning (see once more section 6.6).

In opposition to Prinz' theory (see section 10.3), RM does not define the ***intentionality*** of concepts so narrowly as to require a representational stand-in. Instead, I suggest that the intentionality involved in conceptual abilities takes the form of an intimate qualitative interactivity between agent and environment. The concept 'dog', for instance, has particular properties (because of which it can be or is customarily used in a particular way) because actual dogs have the kinds of properties and stand in the kinds of relations to the rest of the world in ways which constrain and enable appropriate 'dog'-concept-use. Furthermore, there is a kind of social co-construction of conceptual content and reference because of an agent's interaction with other agents, with their own understanding of that particular concept, resulting in the above-mentioned narrative jurisprudence of concept use. This multi-layered constraining-and-enabling dynamic instantiated in the interaction of agent and environment involves what I have called dynamical dimensioned realisation (in section 7.8).

RM allows a greater flexibility of relative 'truth'-preserving concept use, first by relinquishing the use of the term 'truth' and instead using 'appropriateness of use', which denotes a contextually defined pragmatic approach to conceptual meaning (see section 6.11.5). An additional remark is that RM allows for the fact that in many cases, concept use does not involve an intentionally (in the sense of consciously intended) focused referential relationship, but an attempt to adhere to a learned social, cultural, behavioural or linguistic convention, which might not be clearly defined - an attempt, furthermore, which might fail to some extent (e.g. it might not reach the kind of precision that would be required for scientific use) and still cause other people to understand what concept the agent in question intended to use. In such cases, we will notice that our discussion partner is not using a particular word/concept correctly (or at least differently from what we would do or say), but we *think* we know what it is that he means to say or what he is referring to, and we respond as if he had really said what we think he was supposed to say. One might call this approach to truth and concepts 'pragmatic realism': this is what often really happens, and in such cases appears to *work*. In many non-scientific cases, that is good enough.

Finally, Prinz claims that a concept should be individuated in terms of its relations to external entities as well as its interrelatedness with other concepts: a concept should have ***cognitive content*** as well as intentional content. RM satisfies both components of this requirement, but in rather specific ways, especially where it concerns the notion 'content' (see chapter 7). That is, according to RM, a concept is a contentful mental state, but this state is defined in $E_{(i)}C$-terms as an ability to act and react in an agent-environment interaction dynamic - the structure required for this aspect on its own is described in terms of the aforementioned Dynamical Dimensioned Realization, which concerns the intentional connection between concept and world. The second part of the requirement is met in the sense that concept use is justified by referring to myriad other concepts. That is, any concept is always linked to many other concepts, in terms of conceptual abilities being composed of higher-grained conceptual behavioural profiles, and conceptual abilities depending on the implementation of other conceptual abilities to complete a sustained meaningful interaction of agent and environment.

10.6 - In Conclusion

If the RM-model is correct, even if only in principle (e.g. the details of the respective spaces' internal structure are errant to some degree, but the idea of an interaction structure of descriptive domains is on the right track), this can help determine what *form* empirical data within a particular domain (say, behavioural phenomena) needs to have to be applicable to domain-transgressing cases. In other words, experimenters should attempt to translate the data found during some experiment into a description in terms of the contribution of that subdomain to the activities of the agent as a whole, and as embedded in his environment. This could, in principle, constrain or guide the way in which experiments are set up. Furthermore, if the RM would be adopted as a framework describing some structural aspects of the agent's behaviour, hence if the RM was used to inform the presuppositions from which hypotheses are drawn up, this could determine *what kinds* of methodologies and experiments are acceptable at all as proper explorations of cognition. Obviously, scientists working in a particular sub-domain - say, behavioural psychology - would still be free to carry on doing research whichever way they would deem empirically appropriate, but if they wanted their conjectures to be applicable to concepts and cognition in embodied and embedded terms, they *might* need to redesign their experiments. The case described in section 10.4 demonstrates that RM has suggestions to make about the way in which a concept-based curriculum needs to be designed.

For the RM to have such an influence, it needs to be subjected to extensive refinement. The most important task to be carried out concerns the *interpretation of data*: how do we 'translate' empirical results in such a way that they can be encoded in terms of one of the RM's spaces, hence be related to other kinds of data, encoded in the RM's other spaces? Important work in interpreting phenomenal data for use in (neuro-)dynamical models

(Thompson, 2006) might offer some clues regarding how to go about such a task. Another area of the RM-model where there is still a lot of room for improvement involves phenomenology: some suggestions in that direction were made (e.g. about the role of phenomenal experience in the structure of the narratives that lie at the basis of SToCC's inferred accounts, in section 6.6), but the RM-model could use a clearer account of the status it affords to 'what it is like'-judgments.

In the end, the RM is intended to provide a new metaphor for an $E_{(i)}C$-appropriate relate important data-domains (i.e. involving concepts, behaviour, biomechanical properties and environmental physical and social affordances) to eachother. Obviously, this book is merely a sketch, a provisional and hypothetical framework resulting from philosophical analysis, that might have empirical consequences. A lot of work still remains to be done, but I hope the RM helps us gain some headway on the difficult road towards a comprehensive $E_{(i)}C$ theory of concepts and cognition.

= =

[Notes]

Note 1: I agree with Bennett and Hacker's Wittgensteinian inclination (i.e the campaign against hidden Cartesianism) in principle, but I feel their project overshoots its target on two counts: (1) their representation of the way neuroscientists talk, which they use to make their deconstructivist point, is too often sampled from popularizing literature, and in such cases fails to provide an accurate, unbiased portrayal of the *scientific* opinions of these neuroscientists. The kinds of metaphors they criticise do turn up rather often, but this is not always as debilitating as they might claim; (2) they leave very little room for the explanatory or expository use of metaphor, and when dealing with something as conceptually slippery as the mind, sometimes metaphor is all we can use. We should be careful not to discard the good along with all the bad.

Note 2: In this notation, 's' stands for 'situated'; the 'b' - for embedded - was already taken by *embodied*.

Note 3: See section 7.2 for a more thorough look at enactivism, and the criticism Dan Hutto uses to introduce his adaptation, *Radical Enactivism*. Prior to that, section 3.2 will feature a description and critique of an exemplar of the dynamicist variant of $E_{(a)}C$.

Note 4: Dan Hutto suggests 'enculturedness' should be a separate element of the broad 'embodied, embedded, etcetera'-approach. Thompson (2007) defends a similar claim.

Note 5: It is possible to speculate that this kind of negative heuristic - i.e. the development of a multi-tracked, multidisciplinary view with as its main uniting feature the *opposition* to some other view, namely cognitivism - might contribute to the fragmentation and definitional inprecision of the embodied/embedded 'paradigm'. Let's put it this way: when you are constantly busy cutting off the heads of the Hydra with a sword, there is little time left for research and development of higher-tech weaponry - say, tanks and fighter jets. However, the origin of this fragmentation is not an issue that I wish to investigate here; rather, I hope to offer a small contribution towards finding an *antidote*.

Note 6: See section 3.2 for a more thorough (but still brief) discussion of Thelen et al.'s model, and the way their findings are used to construct the 'Radicality Manifold' model. For a lengthier discussion of this model, and the hopes for a dynamicist philosophy of cognition in general, see Van Leeuwen (2005).

Note 7: In classical set theory, an element either is or is not part of a set according to some criterium (or list thereof); this binary conception of category membership yields 0 and 1 as the only possible membership values. Fuzzy set theory, on the other hand, allows for graded category membership: the set of permitted membership values includes, in principle,

0 and 1 plus all real numbers in between. Fuzzy set theory helps provide a formal depiction of the characteristic properties of concepts (as defined by prototype theory) by attributing higher values (i.e. closer to 1) to exemplars that possess more features defined to be typical of some category.

Note 8: Parts of this section were published previously in Van Leeuwen (2005).

Note 9: Dynamical Systems Theory (DST) is concerned with finding mathematical descriptions of the way systems (or aspects thereof) change over time, using differential or difference equations. A dynamical model consists of a *state space*, which is defined in terms of dynamical variables representing the relevant properties of the system, a time set, and an equation or set of equations transforming an initial state of the system at some moment in time into another state at a later time. In case of a continuous time set, this yields a curve in state space expressing the evolution of the variable(s); this curve, consisting of the points the system passes through as it evolves, is called the *trajectory* of the system.

A classical example of a dynamical system is the pendulum, and this system's two-dimensional dynamics (the system's behaviour being described by the pendulum's angle of elevation and its rate of rotation) were already explored by Newton. In another example, in describing Newton's famous falling apple, the relevant dynamical variables would be the apple's velocity and its position.

Dynamical variables that help specify the state of the system in some crucial sense are *order parameters*. Their role can be compared to the parameter 'density' in a model describing the behaviour of a gas: such a parameter represents the macroscopic behaviour of the many individual gas molecules in a compact and efficient manner. Likewise, in DST models order parameters can capture the dynamics of (some aspect of) a complex, inherently high-dimensional system in a low-dimensional fashion. Influential parameters that are somehow external to the system itself are called *control parameters*. In the falling object example, the strength of the gravitational field would be a control parameter.

Depicting behaviour of some system in abstracta, models are often idealised versions of reality and, for instance, ignore friction – a pendulum swinging in a vacuum, without any friction at the hinge point, will retain its amplitude indefinitely. Such systems are called *conservative*, but many *real* systems are *dissipative*, meaning they lose energy, slow down or otherwise tend towards some end state in an asymptotic manner. The point in state space a system starting at some other point A evolves towards eventually is called the limit point of the trajectory through point A. In higher-dimensional systems, a cycle or a torus can also be a limit set of some trajectory. The equilibrium state of the system, i.e. the point or set of points a trajectory tends towards over time, is an *attractor* of the system. The collection of points in state space that a trajectory can start out in to eventually arrive at

(or immeasurably close to) the attractor is the attractor's inset, or *basin of attraction*. The opposite of an attractor - a point (or cycle, and so on) trajectories 'flee' from - is called a *repeller*. A system often has more than one attractor; separating attractor basins are the separatrices. Two possible initial states might be very close together, but if they lie on opposite sides of a separatrix they could end up at very different positions in state space after a while.

Limit sets, in some cases, can be much more complex than points or cycles. Strange attractors are those limit sets that have a very complex geometric shape, and a system exhibiting the behaviour associated with such attractors is said to be chaotic (yet still deterministic, because they are still described by differential equations). It is in these cases that systems exhibit an extreme sensitivity to initial conditions: a minute change in the system's initial state could result in wildly divergent behaviour. Thus, despite the deterministic character of the behaviour of these systems, accurate long-term prediction is often practically impossible because there is always a nonzero measuring inaccuracy that, over time, might throw a spanner of substantial size in the works.

The attractor topology of a system is not necessarily constant over time - in keeping with the sensitivity of chaotic systems to the specifics of initial conditions, small changes in control parameters can cause large shifts in the way a system behaves. Such changes, involving changes in attractor properties or even the sudden disappearance of an attractor (or the emergence of one where there previously was none), are called phase transitions or bifurcations.

Recapitulating, systems described in DST are fully deterministic, but seemingly random phenomena might occur. Non-linear, chaotic dynamical systems are highly sensitive to changes in initial conditions: a system may exhibit fundamentally different behaviour if the initial conditions are modified only slightly. Weather systems, or the eddies, flows and vortices in streaming water are examples of chaotic systems in this sense, and the use of DST to generate models to describe the behaviour of such systems has greatly increased our understanding of such processes. This chaotic behaviour generates special problems regarding the description of such systems using mathematical equations: in systems susceptible to chaotic behaviour, even a miniscule error in the specification of initial conditions while modeling actual systems might render the model inapplicable.

Note 10: The way Thelen et al. construct their model, i.e. in terms of a mathematically defined field, is inspired by several older accounts, amongst them Köhler's field theory and Lewin's topological psychology, [b] interpreted in line with Gestalt theory and behaviourism, and [c] recast in terms of DST.

Note 11: In terms of the 'Radicality Manifold'-model to be developed in this book, Thelen et al.'s model describes how P-space and M-space

collectively constitute B-space - that is, how particular shapes of behavioural dynamics emerge from the interaction of the properties of an agent's body and his environment.

Note 12: A description detailed enough to do justice to the ingenuity of this model is beyond the scope of the current discussion - my objective here is merely to provide a general overview of the model and the successes its creators claim it is capable of achieving, enough to determine its philosophical relevance. For readers desiring a more detailed look, Thelen et al.'s 2001 article is very thorough in its description of the particulars of the model.

Note 13: Note how elements four and five in particular already presuppose the embodiment thesis.

Note 14: The dots above the X intend to denote the order of the temporal derivative. X is a variable expressing distance; an X with one dot is the first-order derivative of X, meaning a certain distance traveled per unit of time, i.e. velocity; an X with two dots is the second-order derivative of X, meaning the change, per unit of time, of the velocity, i.e. acceleration.

Note 15: This generates all kinds of problems involving 'time's arrow', i.e. why time *does* appear to have a distinct direction, despite the lack of support from the mathematical formalism. See Sklar (1977).

Note 16: These hue-cancellation experiments were set up as follows. First, the phenomenally unique hues (that is, say, a yellow devoid of traces of any other hue) would be determined - the corresponding wavelength in nanometres would be recorded. Then, it proved possible to remove the yellowness from a reddish-yellow light (orange) by a light that, seen in isolation, would appear unique blue. The cancellation would consist in all traces of yellowness being gone from the resultant light without there being a hint of blue, i.e. the resultant light would be reddish. After that, the reddishness of the same orange light would be cancelled by a light of unique green. Progressing through the visible spectrum at 10 nm increments and recording all relevant energy levels of cancelling hues, the performances of the red vs. green and yellow vs. blue responses could be determined.

Note 17: I will discuss more ideas by Kimberly Jameson in section 5.2, where her Interpoint Distance Model - an extrapolation of the suggestions put forth in her 1997 article together with D'Andrade - will prove to contain important notions and concepts, to be used in my own account.

Note 18: The 'World Color Survey' is a massive research project intended to improve upon the findings of Berlin and Kay (1969), by charting the ways in which colour phenomenology and colour language are related for many different languages and cultures worldwide.

Note 19: Surface Spectral Reflectance (SSR) is a function specifying what proportion of light an object's surface reflects for every wavelength.

Note 20: The main objectivist theory about colour is *physicalism*, which holds that colour is some kind of physical property. An important argument *in support* of physicalism about colour involves the natural way in which it allows alliances with physics. For public relations purposes amongst most analytically oriented philosophers, at the very least, that is a great advantage. More importantly, with physics in the explanatory toolbox, it is possible to attribute to colours a proper causal role (e.g. it is some type of microphysical structure that causes the reflected light to have a particular wavelength), and to do so with an ontological commitment to relatively few properties. This way, physicalism can accommodate that most basic of intuitions about colour, namely that colour is some property of the object. Hence, on such an account, colour perception can be veridical: an object can actually possess the colour that we perceive it as having.

The main argument *against* physicalism about colour is the problem of *metamerism*: any one perceived colour can be caused by any one out of a disjunct set of Surface Spectral Reflectance profiles. This means that the set of microphysical structures we group together as, say, 'blue-causing' microphysical structures, has nothing in common other than the fact that we, perceivers, *subjectively* perceive them as being the same in some way (namely in terms of apparent colour). An additional problem for physicalism is that the phenomenal properties of colour - for instance, opponency, or the primacy of particular hues (see section 3.2 and chapter 4) - cannot be accounted for in physicalistic terms. There is nothing in the microphysical structures (which cause colour according to physicalism) that stands in the kinds of relations to eachother in the way that, for instance, red and green are opponent colours in perception.

Dispositionalism, involving the claim that colour is the disposition of an object to cause a particular colour sensation, is often classified as an objectivist theory, and technically this is correct. However, this position includes an ontologically puzzling infusion of *subjectivism*. Consider that dispositionalism says a particular object is green just in case said object has the tendency to appear green to normal observers, under normal circumstances. A first glance, this appears to be a clever way of maintaining objectivism (colour is a specific power of an object to appear as such and such), while sidestepping the danger of metamerism that plagues standard physicalism: the definition of a particular colour explicitly includes the perceiver-centric criterion of appearing to be that colour, regardless of the disjunctivity of its physical base.

These apparent advantages come at considerable cost. The first item on the list of disadvantages is that the dispositionalist definition of colour runs the risk of being circular: an intended explanation in terms of a colour being the power of some object to be perceived as possessing a particular colour, does not explain much in an obvious manner. Or if it explains anything at

all, then it does so merely in an indirect manner, because a disposition would need to have a physicalistic base that would provide the actual explanatory power. Furthermore, dispositionalism encounters severe difficulties in defining what 'normal' is in the case of observers as well as perception conditions. In fact, in defining colours at least partly in terms of properties of (the perceptual abilities of) human observers, the vast array of non-human colour-observers is almost automatically disqualified from having worthwhile colour vision under said definition. That is, unless extensive qualifiers are added, which usually do not serve to elucidate the definition. An additional implication would be that a colour would not exist if it went unobserved, and that once again clashes with our intuition that a colour is a property of an object.

Subjectivism about colour, the claim that colour is a mentally efficacious property, most naturally accommodates explanations of the phenomenal properties involved in colour perception mentioned above. The main argument against subjectivist theories about colour is that on such accounts, colour perception commits a global error: objects are perceived as being coloured, while they are not, because subjectivism states that colour is 'in the head', rather than 'on the object'. In everyday perception (i.e. discounting exceptional phenomena such as afterimages), we quite clearly perceive objects (or at least phenomena for which a satisfactory physicalistic description is available, e.g. the sky) as being coloured, and it seems highly counterintuitive to say that what we see as object colours are somehow mental phenomena. Now, as we shall see in sections to come, there are certainly aspects of colour perception that are not in any obvious way linked to objective phenomena, but are instead generated 'internally' - some phenomenal properties of colour experiences (such as the exact specifications of the primary hues) depend quite heavily on the properties of the perceiver's visual system. However, we can quite comfortably say that subjectivism-on-its-own about colour does not do justice to one of our most deep-seated intuitions about colour (apparently supported by findings from phenomenology), namely, that it is usually the object that is (or appears to be) coloured.

Note 21: Section 8.2 contains a closer look at the notion 'affordance'.

Note 22: The following is a cursory description of some of the LMF-model's technical aspects, based on (Wandell 1989). Readers eager to explore the details of this model are encouraged to seek out that article, and the papers mentioned in its bibliography.

The spectral power distribution of the illuminant can be expressed as the sum of the weighted contributions of each basis function:

[Eq. 1] $$E(\lambda_n) = \sum_{i=1}^{N} \varepsilon_i E_i(\lambda_n)$$

$E(\lambda_n)$ = spectral power distibution measured directly; $E_i(\lambda_n)$ = the most efficient basis functions; ε_i = weights of basis functions.

The surface reflectance function at a point can likewise be expressed as the sum of the weighted contributions of each basis function at a specific location on the object's surface:

[Eq. 2] $$S^x(\lambda_n) = \sum_{j=1}^{N} \sigma_j^x S_j(\lambda_n)$$

$S_j(\lambda_n)$ = basis functions; σ^x_j = weight of the *j*-th basis function at spatial position x.

The resultant equation shows the interrelatedness of the ambient light, surface reflectance and receptor sensitivities by defining retinal activity in terms of lighting, reflectance and spectral sensitivity of chromatic receptors:

[Eq. 3] $$\rho_k^x = \sum_{n=1}^{N} R_k(\lambda_n) E(\lambda_n) S_x(\lambda_n)$$

$R_k(\lambda_n)$ = spectral sensitivity (fraction of light absorbed for each wavelength) of photoreceptor of type k; ρ^x_k = number of quantal absorptions of the receptor of type *k* at retinal location *x*.

The perceiving subject's receptor sensitivity is assumed to be constant. Humans have three cone types, with maximum sensitivities at the following wavelengths:

445 nm – 'blue' cone – 'S' (short wave),
535 nm – 'green' cone – 'M' (middle wave)
570 nm – 'red' cone – 'L' (long wave).

The resultant above can be expressed in matrix form, and then simplified:

[Eq. 4] $$\rho^x = \Lambda_E \sigma^x$$

Λ_E = matrix expressing ambient lighting and receptor sensitivity.

Λ_E is constant. In the case of ambient lighting, this is an assumption; therefore, the model is a simplification. If we know the quantum catch at retinal location x, we can compute σ^x, which expresses the reflectance at object location x – the distal chromatic/physical invariant the visual system was supposed to extract from the stimulus.

When the lighting conditions are unknown, Λ_E specifies the linear mapping of an (x)-dimensional surface reflectance representation onto an (x+1)-dimensional receptor response. Properties of the specified plane in (x+1)-dimensional space enable predictions regarding properties of the lighting conditions; the position of a point within that plane enables an estimate of some surface reflectance function.

Note 23: Wachtler et al. (2001) use Independent Component Analysis (ICA). ICA is a decorrelation algorithm for complex data sets: it can find a linear transformation of a vector representation of some signal (for instance, a spectral distribution vector expressing some image) that yields basis functions of the compound signal that are as statistically independent as possible. ICA has no orthogonality constraints (unlike a related decorrelation method, Principal Component Analysis) - this rules out the possibility any orthogonality of basis functions is an artefact of the method rather than a proper representation of the analysed signal.

A useful idea to come out of the computational vision project is that sensory processing serves to reduce redundancy in the information of a scene, thus increasing coding efficiency by decorrelating the component functions encapsulated in the scene's light array. In the case of the inverse optics process (recovering information about an object's surface properties from retinal stimuli), some of the ICA basis functions rather closely resemble natural spectra, raising the possibility that these models might generate hypotheses regarding how biological vision systems might achieve the uncoupling of information about the chromatic aspects of the illuminant from information about the object's surface spectral reflectance (SSR), which would enable colour constancy judgments. The strategy utilised is opponency, which is a rather efficient way of encoding the information both at the retinal and cortical levels. The opponency axes at these two levels do not line up, however, and it remains to be seen how this will affect the phenomenological story of red vs. green and blue vs. yellow. The suggestion by Wachtler et al. (both in their 2001, and in Lee et al. 2002) is that retinal and immediately post-retinal (dLGN) coding along orthogonal opponency axes serves to achieve a reliable transmission of signals from the three chromatic photoreceptor types (decorrelating because of the significant overlap of sensitivity curves of the three cone types), whereas cortical recoding along non-orthogonal opponency axes results in a signal that more accurately reflects the statistical structure of the light coming from the environment.

In Lee et al. (2002), the findings from Wachtler et al. (2001) are extrapolated and fine-tuned. The mechanism of opponency has been claimed to embody an efficient encoding strategy of retinal stimuli. However, Lee et al (2002) suggest opponency arose not merely due to encoding efficiency per se, but also as a reflection of properties of natural spectra: the phenomenon arises as an optimum solution not only in ICA models simulating the sensitivity range overlap of the three human photoreceptor types (which, in an initial hypothesis, would necessitate opponency to decorrelate the signals), but also in hypothetical models where the sensitivity ranges of the receptors show no overlap at all.

Lee et al.'s ICA model yields three main types of basis functions: homogeneous chromatic, oriented achromatic (expressing luminance edges) and colour-opponent (expressing colour edges) basis functions. The chromatic basis functions with the highest contribution represented a light

blue - dark yellow opponency; other components of lesser weight included blue - orange and bluish green - orange opponencies.

The discrepancy between axis orientation between the encoding at the retinal/dLGN level and the cortical level, suggests that the axes of efficient coding don't necessarily align with the cardinal axes of dLGN cells at all levels. At that initial (retinal+dLGN) stage, PCA-type orthogonality emerges as an efficient encoding strategy of the signals received from the three chromatic photoreceptor types, which exhibit significant sensitivity range overlap, but at the cortical level a recoding takes place that results in a much more economical code, that more accurately represents the properties of the light coming from the environment.

It should be noted the ICA basis functions do not show the double opponency of the cells assumed to do much of the coding work in the cortex. Other differences include that the basis functions encoding the red-green opponency mix in contributions from S cones along with the L and M ones. Lee et al. explicitly note the abstractness of their model, and stress the need to take ecological factors (including the *relevance* of the visual signals to the perceivers) into account.

Note 24: Van Hateren and Van der Schaaf mention that the properties, as derived with an ICA model, for which this is the case, are spatial frequency bandwidth, orientation tuning bandwidth, aspect ratio of the receptive field and receptive field length. For the location of the peak of the spatial frequency response, simple cells are much more flexible than the model's predictions would suggest.

Note 25: There is a conflict here between Thompson's enactivism and Shepard's representationalism. Later, in chapter 7, I will have more to say about how I intend to fit the notion representation into an $E_{(i)}C$-appropriate account.

Note 26: Effectivity is the animal-centered counterpart of the object-centered affordance. Turvey, Shaw, Reed and Mace (1981) say: 'the dispositions of an organism-free world and the dispositions of an organism-populated world (...) are not of the same order. The latter are ontologically condensed out of the former, so to speak, by the presence of living things.'

Note 27: See later, in section 5.2 and note 31, when I will define this as the 'Neurophysiological Yield'

Note 28: A possible counter-argument can revolve around the rejection of the Sapir-Whorf-thesis: if the structure of language has no significant effect on the structure of perception (and cognition), these linkages the other way around, namely between perceptual space and concepts, might also be less secure. But all this does is remove the necessity-aspect: in that case, certain perceptual (or, dare I say it, *phenomenal*) features are no longer necessarily linked to certain non-perceptual content, or content from a

different modality. But the idea that percepts have no isolated existence, but rather appear in context, remains: a colour is always a colour *of* something (say, an object), and we often use that colour as an indicator of some other property (redness of an apple means it is ripe, redness of a traffic light means crossing the street at that very moment is dangerous, and so on). In other words, the integration of perceptual space into the lowest reaches of conceptual space does not result in one-on-one mappings of percepts and concepts, but there are definite correlations and co-occurrences. What co-occurs with what can be left to be determined by context and creativity.

Note 29: This method of expanding perceptual space into a conceptual space resembles Gärdenfors' (2000) project, so obviously I am quite sympathetic to some of his claims, but I feel his account misses the mark in some nontrivial spots. See section 10.2 for a more thorough discussion of the similarities of and differences between SToCC (plus its expansion, the Radicality Manifold, to be developed later in this book) and Gärdenfors' theory.

Note 30: Retinal tetrachromats might have four kinds of chromatic receptors ('cones', of which normal humans have three kinds) on their retinas, but lack the fully developed neural processing capacity to actually see colours in a way that captures this extra information. Perceptual tetrachromats do have the capacity to see additional hues and hue mixtures.

Note 31: I define the 'neurophysiological yield' as the prestructured space of hue / brightness / saturation saliencies that is the result of the workings of the receptors on the retina and the subsequent neural processing of the stimuli, in accordance with their physical specifications. Put in the kinds of pastoral and nostalgic terms that the $E_{(i)}C$-programmes would have us rescind (and hence, that my project is supposed to offer an alternative to), the neurophysiological yield can be thought of as a pre-cognitive presentation of the neuro-physiological processing of colour signals. So: the neurophysiological yield defines a structure of hue foci in terms of salience gradients, but this physiologically determined structure need not coincide exactly with the properties of some individual's perceptual colour space.

Note 32: Bernard Harrison (1973) suggests that colour should be defined *operationally:* "(...) 'colour' corresponds not to a *thing* (an object of reference) but to an operation. (...) (T)he individuation of the fundamental modalities of perception depends upon the fact that the items falling under these modalities exhibit continuous systems of internal relationships." (Harrison 1973, pg. 87)

These internal relationships are not due to the properties of a priori colour experience in isolation, but rather emerge in the learning and using of colour language, in the person's operating within the constraints set by the way colour names are used in his language community. In other words, Harrison claims, calling an object 'yellow' is not to say that it somehow

coincides with some pregiven point in perceptual space, but rather that it fits somewhere within this language-based constellation of relationships. A colour, being a relational entity, cannot exist divorced from this embeddedness in its relational web.

Note 33: This involves a *syntactic* conception of information (i.e. the way information is defined in communications technology, namely as a structural feature of signals, e.g. Wiener 1961), combined with an agnosticism about the origin of this information.

Note 34: This involves a *semantic* conception of information (a red traffic light means 'stop'), with pragmatic implications.

Note 35: This claim is somewhat similar to the one professed by Maund (1981, 1995), and in fact inspired by it. See section 6.10 for more on Maund's ideas, and the way in which my suggestion differs.

Note 36: This remark demonstrates that, for the moment, I wish to put aside one of the most important problems of the philosophy of cognition, namely the problem of accounting for the 'what-it-is-like'-ness of phenomenal experience: obviously, the chemical properties of H_2O-molecules do not explain what it is like for me to feel wetness. This issue is not wholly irrelevant to the topic at hand: phenomenology was listed above as one of several disciplines that co-inform the complex concept 'colour', and it will be mentioned again. For instance, the 'Radicality Manifold' model to be developed later in this book is intended to be phenomenally appropriate. However, I would like to submit that the relevance of phenomenology to some concept's explanation is not sufficient for that concept to be complex: true complexity requires the array of incompatible descriptive strategies to be broader and more textured. If this condition were to be dropped, the concept 'water', claimed to be non-complex above, would also be complex (it would not be possible to capture the phenomenal experience associated with drinking cool water on a hot day in physical or chemical terms, for instance), and so would many, many others. This would hollow out the notion of conceptual complexity. So, to recap: the relevance of phenomenal content as a partial description of some concept's meaning *does* contribute towards that concept being complex, but it is not in itself a sufficient condition.

Note 37: This claim implies a notion of normativity. The connection between concepts and normativity will be addressed later, in section 9.2.

Note 38: For a while, I toyed with the idea of relinquishing the use of the term 'concepts', and instead use either 'cogcepts', because concepts in the account to be developed here form interlocking and co-dependent 'cogs' in a larger structure, or 'defcons', which is short for 'deflationary concepts'. I decided to go with the old and trusted term 'concepts' anyway, but with the cautionary note about the deflationary character reproduced above, and forget about 'cogcepts' (too awkward) and 'defcons' (too cute). Another

(important) reason to retain the familiar word is because concepts in the classical meaning really are merely limit cases of a much broader spectrum, which is a hypothesis that - I hope - will appear acceptable to the reader once he has finished this book.

Note 39: This should also be taken to mean that the claims expressed in SToCC are at odds with those of, for instance, Bermudez (1998), who draws a clear boundary between conceptual and nonconceptual content, the former requiring linguistic abilities. It is my belief that the use of such a criterium denies too many exceedingly clever, but nonlinguistic animals the possession of concepts, which appears a highly anthropocentric thing to do. That in itself is not an argument, but a good case can be made for the claim that the abilities of (for instance) New Caledonian crows, chimpanzees and bottlenose dolphins show nontrivial overlap with humans exactly in those areas where some form of concept-use would be in play, i.e. social and/or tool-involving abilities. I hope the remainder of this book will make it clear how the broadening and diversification of the 'concept' concept provides us with a better way of relating conceptual action and knowledge to other kinds of action and cognition, also as it is displayed by non-human animals.

Note 40: All the claims made so far might raise questions about the status of concepts in SToCC: are they real, and are they even *representations* of any kind, as they are in many other theories of concepts? The answer is that in SToCC, concepts are real, but they don't need to be representations for that to be the case. They can certainly be represented or representations in some cases, for instance when they perform a role in a 'representation-hungry problem' (Clark 1997). This will happen when we utilise a concept (or meaningful constellation of concepts) to formulate a hypothesis or prediction in the absence of the conceptualised entities. However, in most cases, concepts, understood as dispositions towards acting in such and a such a way in a particular kind of situation, will not be representations of the classic kind. But even in that latter class of cases, concepts are real because behaviour (be it cognition, action or locution) is real. Perhaps it helps if we borrow a bit of vernacular from Varela and colleagues (see for instance Noë, 2004) by saying that we don't *use* concepts - for that would mean they are, in essence, entities with specific properties, things that we can grab hold of or process in some way -, but that we *enact* concepts. A concept is a structural aspect of what we do and say and think, rather than an independent entity. In that sense, concepts are as real as it gets, despite also being quite ephemeral.

Note 41: As a preliminary warning, I should stress that SToCC does *not* claim that the formation of a specific colour concept consists of 'mental coordinate calculations' in said space, except as a most abstract description of the process of concept formation - so abstract, in fact, to be positively misleading.

Note 42: It is important to realise that following this trail towards the explanatory definitions of a concept's associated theory (theories) is not

necessarily the same as moving towards the non-conceptual basis of conceptual space - see section 6.9 for more on this.

Note 43: Although, of course, the difference with real fractals is that the structures at a 'lower' level of conceptual space need not be similar to any structure at the 'higher' levels.

Note 44: A more in-depth discussion of the differences between SToCC and Theory theory follows in section 6.11.3.

Note 45: See sections 6.7 and 6.8 for more on being able to use the same concepts, despite minute differences in inferred accounts.

Note 46: Please note that this move takes us away from the science-centric explication of the 'colour'-concept as featured in chapter 4, as well as sections 6.1 and 6.2. The introduction of the 'inferred account' as a way to specify the meaning of a concept is the main component that helps transform this earlier discussion into an account about concepts in general, namely SToCC.

Note 47: There is a relatively innocuous sense in which mental phenomena should be regarded as being diachronic. This sense concerns the fact the formation of mental states (sensations, thoughts, and so on) is a process that requires a certain timespan to take place. For instance, given a particular modality as activated in a specific way (e.g. the sensation of feeling as evoked by a stimulus of a particular intensity at a specific location of the skin, having certain sensitivity properties), stimuli need to persist for a minimum amount of time for a conscious sensation of the stimulus to be formed (time is needed, partly for the associated activity in various relevant brain regions to achieve neural integration; see Tononi, 2004). However, the diachronicity of mental states in a semantic sense is at least partly independent from this point, for it depends on causally efficacious connections that are *discontinuous* in time. For instance, a memory of a past event, and therefore at least some aspects of the past event itself, can exert significant influence on an occurrent decision (Slors 2001).

Note 48: For instance: in the perceptual colour space example, there is a structure of focal colours - the best examples of various hues - which is, to a large extent, determined by the properties and relations established in the neurophysiological processing of colour stimuli.

Note 49: This description implies that an *enslaver* is, in some sense, similar to Jesse Prinz' (2002) *proxytype*. The latter is a representation in working memory (a simulation process - be it uni- or multimodal - of the kind involved in actually encountering the entity that is tokened by the concept) that stands in for a more elaborate account of what a concept is supposed to denote. However, the most important difference is that an enslaver is not a mental representation. Rather, it is a constituent of conceptual space, which offers a *descriptive* account of $E_{(i)}C$-behaviour rather than of an

agent's internal (mental) content (but see chapter 7 for ideas about how representation can fit into an $E_{(i)}$C-apropriate theory of concepts, and section 10.3 for a more detailed discussion of Prinz' Proxytype theory).

Note 50: Suppose, for the sake of argument, that 'divine retribution' constitutes a complete description of the explanation used at the time, and that this argument did not involve further reference to bacterial infection as a mechanism utilised by whatever divine power was invoked.

Note 51: See section 6.11.1 for additional examples of such 'same concept/different inferred account'-cases, and a description of the role of enslavers in solving the inherent instability of the 'concept'-concept.

Note 52: Minor integer-substitution is mine: '(25)' became '(1)', '(26)' became '(2)'.

Note 53: The example above, of sickness at different granularities resulting in two distinctly different *kinds* of descriptions, *might* be like this.

Note 54: The idea, developed in the philosophical tradition of e.g. Henri Bergson, is that in experiencing or remembering events from the past, meaningful occurrences are differently indexed than periods that are light on personally significant events. In experience, eventful periods seem to progress quickly ('time flies when you're having fun'), whereas boring periods can appear to last much longer than they actually do. In memory, the reverse appears to be the case: more is retained from high-density periods than from empty, eventless time.

Note 55: The activity of the sensorimotor system (e.g. premotor area F4 in the macaque brain) serves to integrate stimuli from different sensory modalities, and this integrative effect consists of *action simulation* - an immediate reaction of simulating (with the purpose of planning for) a potential (re)action when some sensory stimulus(-cluster) is present. Understanding a meaningful action (presumably in conceptual terms) involves using the same neuronal regions that are activated in imagining, and imagining such an action is closely linked to the kind of brain activity that occurs while actually performing that action, or seeing some other human doing so. These processes require the kind of multimodal integration offered by the sensorimotor system, so claim Gallese and Lakoff. This way of using the sensorimotor system combines Lakoff's earlier theory on the way higher cognition involves a metaphorical transformation of lower-cognitive processes and activities (Lakoff and Johnson 1980, 1999), and the work on mirror neurons pioneered by Rizzolatti, Gallese and their colleagues (Gallese et al. 1996, Rizzolatti et al. 1996).

Note 56: The kinds of schemata invoked by Gallese and Lakoff describe the occurrent neural activity associated with actually performing the relevant bodily act, seeing someone else perform such an act (i.e. involving the activation of the so-called ‘mirror neurons’), *and* during acts of mentally

simulating (some aspect of) performing the act (e.g. thinking about doing so). This last kind of schema activation is important for the link of sensorimotor schemata with cognitive concepts that are not (or not in an immediately obvious way) connected to concrete bodily acts.

Note 57: I probably need to implement an explication of vernaculars here: earlier (section 6.2) the shift from the superposed concept to a conception was characterised as a reduction with remainder. However, this is clearly not a reduction of complexity, but rather a reduction of a more generally applicable concept to a conception that is higher in detail, but more constrained in terms of its domain of application, and using notions and explanatory accounts relevant to 'a more basic level of reality' in mereological terms, e.g. microphysics.

Note 58: See section 7.2 for qualifications of the notion 'content' in this context.

Note 59: In Vantage Theory, category construction (i.e. establishing a structure, upon which inclusion decisions can be made) involves a push-pull-system of similarity and difference judgments (for these are reciprocals), relative to the fixed coordinate of a hue focus: the balance of similarity and distinctiveness judgments defines internal structure of category. The process occurs in multiple stages: for example, the initial fixed coordinate R ('R' for perfect red, i.e. the hue focus which forms a reference point) and an initial emphasis on Similarity (mobile coordinate S) define the range surrounding the focus that might still be sufficiently similar to be called by the colour term associated with the focal point. In the second stage the inherently mobile S is treated as a fixed coordinate, and D, a difference judgment, is introduced as a mobile coordinate; this level-shift is called 'zooming in'. MacLaury constructs Vantage Theory as metaphorically similar to utilising space-time coordinates, so this 'zooming in' might be akin to picking S as some characteristic point in a new inertial frame. The job of D is to stop the S operator from extending the range indefinitely, and this occurs at those locations in colour space which exhibit sufficient difference from the hue focus to be categorised in a different category. Thus, D defines the width of the category associated with the hue focus R. This way, a subset of colour space is defined as a category with a specific internal structure, with values of category membership (relative to the focus) that might be attributed to locations within that category.

Levels	Fixed Coordinates	Mobile Coordinates	Entailments
1	R	S	focus, range
2	S	D	breadth, margin

[Figure 27: model of the 'red' category in Vantage Theory; adapted from MacLaury 2002]

The account is named Vantage Theory because MacLaury intends the categorization mechanism described by this account as involving the subject occupying a specific 'location' within colour space and establishing a category from there, analogous to the vantage point - say, a specific spot at a certain distance from various landmarks, spatially as wel as temporally (as a function of motion) – from which a person constructs space-time (i.e. an apprehension of the contents of the spatio-temporal array surrounding him). The construction of colour-categories in this way is unconscious, says MacLaury. The construction of a category then involves shifting attention around past various aspects of the frame of R, S and D; only one vantage can thus be occupied at a time, hence a focus on S as the fixed point and its relation to the mobile coordinate D precludes attributing attention to the similarity relation for a given coordinate and the hue focus. Still, in attending to one aspect, the rest of the structure (the 'zooming hierarchy') is already present, at least implicitly.

There can be dominant and recessive vantages, for instance in the case of the Hungarian *piros* and *vörös*, both denoting a kind of red, the former being more general, much like the regular 'red' in English, the latter a less often used, but much more specific (namely with a connotation of intensity and/or passion) and semantically versatile term, often found in poetry. The dominant vantage exhibits a greater emphasis on *similarity* between stimuli, hence *contracting* the distance between reference point and outlying coordinates, and in general being a more coarse-grained process. A recessive vantage, on the other hand, exhibits a tighter focus on *difference*, fosters a *widening* of the distance between the viewpoint and coordinates, and allows for a greater specificity, objectivity and analiticity in judgments. This way, Vantage Theory offers a model to predict both the broader scope of 'piros' and the greater semantic depth of 'vörös'.

Note 60: However, two major differences with Vantage Theory lie in the claims that (1) the similarity-difference push-pull-system in SToCC is not exclusively cognitive in origin or effectivity range, and that (2) the categories, concepts and enslavers posited by SToCC are not necessarily representations. What all the theoretical notions from SToCC do, is provide a framework for *describing behaviour* (cognition, action and locution), and this postulate explains both differences. That is, developments and processes in not just conceptual space, but the agent as a whole and his environment contribute to the ongoing process of categorization and concept formation. This is not to say that concept formation, in SToCC, has nothing to do with cognition; quite the opposite, in fact. However, in SToCC, even cognitive phenomena cannot ultimately be thought separate from the interaction dynamic involving all four spaces described by the 'Radicality Manifold'-model to be introduced in chapter 8. Furthermore, only a subsection of the kind of dispositions 'encoded' in conceptual space involves representations, and this representational zone is firmly embedded in a broader dynamic (i.e. of an agent in a world).

Note 61: In the 'Phaedrus' (265d-266a), Plato describes Socrates as having said the following: "The second principle is that of division into species according to the natural formation, where the joint is, not breaking any part as a bad carver might."

Note 62: In section 6.7, the example given deals specifically with the concept 'justice'. The arguments provided there serve as a justification of the claim made here, about the power of enslavers to sidestep the 'missing prototypes'-problem.

Note 63: Richard Dawkins (1976; 1982) (in-)famously posited the *meme* as a cultural replicator, analogous to the gene as a genetic replicator. Daniel Dennett (1991; 1995), almost as famously, picked up on this idea. He describes memes as 'complex ideas that form themselves into *distinct memorable units*' (Dennett 1995, pg. 344, emphasis his).

Note 64: An important aspect of *colour* can be characterised in these terms: one of the roles that colour performs is an ecologically an evolutionarily significant way of 'quick 'n dirty' information transfer. That is, colour can serve as an indicator of a complexly realized, less visible property, such as the ripeness of fruit or the toxicity of an insect.

Note 65: Given the suggestion that the contents of a concept, and the ways in which the concept are intended to apply, are so fundamentally context-dependent (e.g. applicable at a particular granularity, and dependent on a potentially idiosyncratic narrative account - see section 6.7), speaking of *truth*-conditions is unwarranted. This is because doing so would burden even casually utilised, disposable or time-and-place-locked (sub-)concepts with a responsibility to meet rigid formal rules. Instead, we can speak of *appropriateness-conditions* or *appropriateness-of-use-conditions*, incorporating the following two aspects: (1) the explanatory potential of the best available version of the justificatory account associated with the concept, which would include the measure of fit with the data, generativity of novel predictions, and so on; and (2) the sophistication of the concept as compared to said account, in terms of completeness and measure of adherence to the implications of the account.

Note 66: However, see chapter 7 for a closer look at the role of representation in StoCC in preparation for the extrapolation of SToCC, the 'Radicality Manifold'-model.

Note 67: See section 9.1 for more on SToCC's role in the emergence of meaning: several pieces of this puzzle, the 'Radicality Manifold'-model and its implications for the notion 'content' in particular, have yet to be provided.

Note 68: Biosemiotics is biosemantics without the semantics: the 'states' involved have no truthvalues.

Note 69: Parts of this section were published previously in Van Leeuwen (2005).

Note 70: A Fregean Thought can be characterized as the sense (*Sinn*) of a sentence (e.g. the sense of 'The Eiffel Tower is in Paris').

Note 71: Thanks to Tjeerd van de Laar, whose use of Gillett's theory in his dissertation helped me realise its virtues.

Note 72: Note that these different kinds of properties ('properties of the [1] physical and [2] social environmental processes that the agent is immersed in, as well as [3] the biomechanical properties of his own body') align neatly with the different kinds of properties as expressed in figures 6 and 7, which depict the various factors that influence colour perception.

Note 73: Taking into account a purported distinction between intentional behaviour and mere bodily movement, consider this additional scenario: do we respond differently to an adult and cognitively fully functional human who inadvertantly walks into a vase, thus knocking it over, than to the same person when he shoves the vase in a premeditated fashion? I would say yes. When called upon to account for what happened, the explanation the culprit gives in the former case might something like 'It wasn't me, *I* did not do this'.

Note 74: At one time during the writing of this book I was working in my back yard, which had not been tended to for a while, removing patches of ground-elder, also known as Bishop's weed (Aegopodium Podagraria). This plant is extraordinarily difficult to eradicate, because it produces vast and intricate root systems, and can re-grow quickly from even the smallest leftover root fragment. The weed was annoyingly difficult to get rid of, but its underground interconnectedness did serve as an inspirational metaphor for the kind of structure I suggest is also present in conceptual space: concepts and components of concepts depending on other concepts and concept-components via vast networks of inferential 'roots'. Hence the name 'Radicality Manifold': 'radix' is Latin for 'root', and conceptual space is described in terms of a mathematical space (manifold) with particular 'root-network-like' properties. It is this conceptual space that will be expanded into a broader framework about embodied/embedded cognition in general.

Note 75: See section 9.1 for more on what SToCC/RM has to say about the role of concepts in establishing meaning.

Note 76: An oft-mentioned example of this occurs in Rayleigh-Bénard systems: convection rolls are large-scale dynamic structures that constrain the movements of the liquid's constituting molecules. See Kelso (1995).

Note 77: Thanks to Marc Slors for forcing me to be more explicit about this.

Note 78: Note that this is a different use of the term 'functional cluster' than found in Gallese and Lakoff (2005), a paper referenced in section 6.9.

Note 79: This is not (or at least not necessarily) the arena where the classic problem of mental causation arises. C-space is not the mind, but rather the dispositional space showing the structure of possible dispositions towards exhibiting behaviour (including cognitive behaviour), and B-space is not *just* the movement of bodyparts, but also includes locution and cognition. This is to express the notion that the notion of 'mind' is utterly useless without taking into account what kind of action is possible because of this mind, and especially that the body is essential in determining what these actions are. A disembodied mind is quite literally nothing, and nothing is explained by invoking it.

Note 80: As such, Gallagher offers a non-cognitivistic alternative to the two standard approaches in the 'theory of mind'-debate, namely simulation theory and theory theory (the latter being a theory that is distinct from the Theory Theory of Concepts, discussed in section 6.3.3 above).

Note 81: Besides *affect attunement*, Stern also uses the elegant notion *affect contagion*.

Note 82: Many thanks to Fred Hasselman for introducing me to this discussion.

Note 83: Chemero and Turvey (2007) also discuss Vera and Simon (1993), which offers a computational, representational variant of the affordance-concept, defining an affordance as a representation which enables an efficient mapping between a representation of what the world is like, and a representation of the kinds of action the agent should or could perform. I would have to agree with what Chemero and Turvey imply, namely that this view is not very $E_{(i)}C$-compatible.

Note 84: See section 8.7.2 below for a closer look at the similarities, but especially the differences between Gärdenfors' theory and my ideas.

Note 85: Natural properties are properties that are most significant to an evolved animal's survival, and are - parallel with Quine's (1969) definition of the notion 'natural kind' - *projectible*, which means that they support inductive reasoning. 'Convexity' means that if points x and y are elements of some subset of conceptual space, all points lying between x and y also belong to that subset. Gärdenfors now claims that considerations of cognitive economy suggest that natural properties, if defined in terms of regions of conceptual space, should be *convex* regions: it would make little evolutionary sense to suppose that such important properties would correspond to irregularly-shaped regions, for in that case understanding the coherence of different tokens of the property (to be expressed as different locations in such an irregularly shaped region) and memorising these properties would require a lot of mental processing. This does not appear to

be what is suggested by experimental results involving the understanding of natural properties.

Note 86: Putnam's (1975) *Twin Earth* thought experiment served to pump the intuition that two identical agents with identical (down to the very last sub-atomic particle) internal states, might still be in different *mental* states, because those internal states refer to different external states: when the resident of Earth thinks of the substance 'water' (a clear potable liquid which, under specific atmospheric conditions, freezes at 0°C and boils at 100°C), he is referring to a substance with molecular formula 'H_2O', whereas when the resident of Twin Earth thinks of the substance 'water' (a clear potable liquid which, under specific atmospheric conditions, freezes at 0°C and boils at 100°C), he is referring to a substance with molecular formula 'XYZ'.

Note 87: This second argument is basically the claim that a brain in a vat is impossible.

Note 88: There is a sizeable subset of philosophy of cognition and psychology that is engaged in finding proper descriptions of our use of these 'theory of mind' abilities (Goldman 2006, Gallagher 2005).

Note 89: Or perhaps this means that what we mean when we talk about 'the mind' is what needs to be reconsidered. That, of course, is the very core of the $E_{(i)}C$-project.

Note 90: Of course, if there had been a proper match, RM would have been the derivative theory, because Prinz' book came first. However, RM was developed on the basis of an extrapolation of a theory of embodied/embedded colour perception and 'colour cognition'; the comparison with Prinz' theory was made *after* I came up with 'enslavers' and such.

Note 91: For instance: Latin is, purportedly, a dead language. Still, many thousands of young people enroll in grammar school (or comparable programs) each year all throughout the world. The justification is that studying Latin trains highly useful cognitive abilities, and introduces the student to many ideas, texts and works of art that have had a strong influence upon the formation of Western thought and culture. In that sense, Latin being obsolete as a language does not necessarily make it obsolete as a subject of study. For another example, consider studying substance or property dualism as positions in the philosophy of mind, to understand what is new and different about $E_{(i)}C$. In this sense, studying outdated ideas can help one get a better grip on the ideas that were developed as alternatives to those now-rejected theories.

[Bibliographical References]

Alvarado, N. and Jameson, K. A. (2002), 'The use of Modifying Terms in the Naming and Categorization of Color Appearances', *The Journal of Cognition & Culture*, vol. 2, issue 1, pg. 53-80.

Armstrong, S., Gleitman, L. and Gleitman, H. (1983), 'What Some Concepts Might Not Be', *Cognition* 13: 263 - 308.

Bennett, M.R. and Hacker, P.M.S. (2003), Philosophical Foundations of Neuroscience, Blackwell Publishers, Oxford / Malden, MA.

Berlin, B. and Kay, P. (1969), 'Basic Color Terms - Their Universality And Evolution', Berkeley / Los Angeles: University Of California Press.

Bermúdez, J. L. (1998), 'The paradox of self-consciousness', The MIT Press, Cambridge, MA.

Birbili, M. (2007), 'Making the Case for a Conceptually Based Curriculum in Early Childhood Education', *Early Childhood Education Journal*, Vol.35, No.2, pp 141-147.

Bosman, A.M.T. (2008), 'Pedagogische Wetenschap: Koorddansen tussen Kunst en Kunde', Uitgeverij Eenmalig, Hilversum.

Bosman, A.M.T. and Schraven, J.L.M. (2008), 'Zo Leer Je Kinderen Lezen en Spellen in het Speciaal Basisonderwijs', Tijdschrift voor Remedial Teaching, Vol. 16, No. 1, pp. 26-29.

Block, N. (1998), 'Conceptual Role Semantics', from 'The Routledge Encylopedia of Philosophy', E. Craig (Ed.), London: Routledge. [URL:http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/ ConceptualRoleSemantics.html]

Block, N. (2003) Do Causal Powers Drain Away? Philosophy and Phenomenological Research, vol. LXVII, no. 1, 133-150

Boynton, R.M. and Olsen, C. X. (1987), 'Locating Basic Colours In The OSA Space', *Color Research And Application*, 12: 94-105.

Brandom, R. B. (1994), 'Making It Explicit: Reasoning, Representing, and Discursive Commitment', Harvard University Press, Cambridge, MA / London.

Brandom, R. B. (2000), 'Articulating Reasons: An Introduction to Inferentialism', Harvard University Press, Cambridge, MA / London.

Bransen, J. (2000), 'Normativity as the Key to Objectivity: An exploration of Robert Brandom's Articulating Reasons', *Inquiry*, Vol. 45(3), pg. 373-392.

Brentano, F. (1874), 'Psychologie vom empirischen Standpunkt', Duncke and Humbolt, Leipzig; English translation: 'Psychology from an Empirical Standpoint' (1973), translated by Rancurello, A.C.; Terrell, D.B.; and McAlister, L., Routledge, London.

Byrne, A. (2005), Perception and Conceptual Content, In *Contemporary Debates in Epistemolog*y, eds. E. Sosa and M. Steup, Publishers, Oxford / Malden, MA.

Byrne, A., and Hilbert, D.R. (2003), 'Color Realism and Color Science', *Behavioral and Brain Sciences* 26, pp. 3-64 (target article with commentaries).

Carver, L.J. and Bauer, P.J. (1999), 'When the Event is more than the Sum of its Parts: Nine-Month-Olds' Long-Term Ordered Recall', *Memory* 7, pp. 147-174.
Chaitin, G.J. (1999), 'The Unknowable', Springer-Verlag, Berlin / Heidelberg / New York.
Chemero, A. (2003). An Outline of a Theory of Affordances. *Ecological Psychology*, Vol. 15, pg. 181-195.
Chemero, A. and Turvey, M.T. (2007), 'Complexity, Hypersets, and the Ecological Perspective on Perception-Action', *Biological Theory*, Vol. 2, No. 1, pg. 23-36.
Churchland, P.M. (1995), 'The engine of reason, the seat of the soul : a philosophical journey into the brain', The MIT Press, Cambridge, MA.
Clark, A. (1997), 'Being There: Putting Brain, Body, and World Together Again', The MIT Press, Cambridge, MA.
Clark, A. and Chalmers, D. (1998), 'The Extended Mind', *Analysis* 58, pp. 7-19.
Crane, T. (2003), 'The Intentional Structure of Consciousness', in: *Consciousness: New Philosophical Perspectives*, Q. Smith and A. Jokic (eds), pp. 33-56, Oxford University Press, Oxford.
Croft, W. (1991), 'Syntactic Categories and Grammatical Relations - The Cognitive Organization of Information', The University of Chicago Press, Chicago / London.
Damasio, A. (1994), 'Descartes' Error: Emotion, Reason and the Human Brain', Putnam, New York, NY.
Damasio, A.R. (1999), ' The feeling of what happens : body and emotion in the making of consciousness', Harcourt Brace, New York.
Dawkins, R. (1976), 'The Selfish Gene', Oxford University Press, Oxford.
Dawkins, R. (1982), 'The Extended Phenotype', Oxford University Press, New York
DeCock, L. (2006), 'A physicalist reinterpretion of 'phenomenal' spaces', *Phenomenology and the Cognitive Sciences*, 5, pp. 197-225.
De Jaegher, H., and Di Paolo, E. A. (2007), 'Participatory sense-making: An enactive approach to social cognition', *Phenomenology and the Cognitive Sciences*, 6(4), pp. 485 - 507.
Delahunt, P. B. and Brainard, D. H. (2003), 'Does human color constancy incorporate the statistical regularity of natural daylight?', *Journal of Vision* ref: JOV-00066-2003
Dennett, D.C. (1991), 'Consciousness Explained', Little, Brown (Boston)
Dennett, D.C. (1995), 'Darwin's Dangerous Idea', Simon & Schuster, New York
Dooremalen, H. (2003), 'Evolution's shorthand : a presentational theory of the phenomenal mind', dissertation Radboud University Nijmegen.
Dretske, F. (1988), 'Explaining Behavior: Reasons in a World of Causes', The MIT Press, Cambridge, MA.
Eliasmith, C. (1995), 'Mind as a Dynamical System', Master's Thesis, University of Waterloo, Ontario, Canada.
Eliasmith, C. (2003), 'Moving Beyond Metaphors: Understanding the Mind for What it is ', *The Journal of Philosophy* Vol.C, 10, pp. 493-520.

Evans, G. (1982), 'The Varieties of Reference', John McDowell (ed.). Oxford University Press, Oxford.
Fodor, J.A. (1975), 'The Language of Thought', Harvard University Press.
Fodor, J.A. (1981), 'The Present Status of the Innateness Controversy', in *Representations: Philosophical Essays on the Foundations of Cognitive Science*, pp. 257-316, The MIT Press, Cambridge, MA.
Fodor, J.A. (2004), 'Having concepts: A brief refutation of the twentieth century', *Mind and Language* 19, pp. 29-47.
Fodor, J.A. and Pylyshyn, Z. W. (1981), 'How direct is visual perception? Some reflections on Gibson's "Ecological Approach" ', *Cognition* 9: 139-196.
Gallagher, S. (2005), 'How the body shapes the mind', Clarendon Press, Oxford.
Gallagher, S. and Hutto, D. 2008. Understanding others through primary interaction and narrative practice. In: Zlatev, Racine, Sinha and Itkonen (eds). *The Shared Mind: Perspectives on Intersubjectivity* (17-38). Amsterdam: John Benjamins.
Gallese, V.; Fadiga, L.; Fogassi, L. and Rizzolatti, G. (1996), 'Action recognition in the premotor cortex', *Brain* 119: pp. 593-609
Gallese, V. and Lakoff, G. (2005), 'The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge', *Cognitive Neuropsychology* 21.
Gärdenfors, P. (2000), 'Conceptual Space: the geometry of thought', The MIT press, Cambridge, MA.
Gibson, J.J. (1979), 'The Ecological Approach To Visual Perception', Boston: Houghton Mifflin
Gillet, C. (2002), 'The Dimensions Of Realization: A Critique of the Standard View', *Analysis* 62, pp. 316-323.
Goldman, A.I. (2006), 'Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading', Oxford University Press, New York.
Gopnik, A., and H. Wellman (1994), 'The theory theory'. In: 'Mapping the mind: Domain specificity in cognition and culture', edited by Susan A. Gelman and Lawrence A. Hirschfeld. Cambridge: Cambridge University Press.
Guttenplan, S. (ed.) (1994), 'A Companion to the Philosophy of Mind', Blackwell Publishers, Oxford/Malden, MA.
Hardin, C. L. (1988 / 1993 [revised]), 'Color For Philosophers - Unweaving The Rainbow', Indianapolis: Hackett.
Harman, G. (1998), '(Nonsolipsistic) Conceptual Role Semantics'; [URL:http://www.nyu.edu/gsas/dept/philo/courses/concepts/NonSolips.html]
Harrison, B. (1973), 'Form and Content', Basil Blackwell, Oxford.
Haugeland, J. (1991), 'Representational Genera', in: 'Philosophy and Connectionist Theory', W. Ramsey et al. (eds.), Erlbaum.
Hobbs, J.R. (1985), 'Granularity', *Proceedings of the Ninth International Joint Conference on Artificial Intelligence* 432-435, San Mateo, CA: Morgan Kaufman.
Hofstadter, D. (1979), 'Gödel, Escher, Bach: an Eternal Golden Braid', Basic Books, New York.

Hofstadter, D. (2007), 'I Am a Strange Loop', Basic Books, New York.
Hubey, H.M. (1997), 'Logic, physics, physiology, and topology of color', commentary on Saunders and Van Brakel (1997) - 'Are there nontrivial constraints on colour categorization?', Behavioral And Brain Sciences, 20: 167-228
Hurvich, L.M., and Jameson, D. (1957), 'An opponent process theory of color vision', *Psychological Review* 64, pp. 384-404
Hutto, D. (2006), 'Unprincipled Engagements: Emotional Experience, Expression and Response', in: Richard Menary (ed.), 'Radical Enactivism: Intentionality, Phenomenology and Narrative - Focus on the Philosophy of Daniel D. Hutto', John Benjamins Publishing, Amsterdam.
Hutto, D. (2007), 'Folk Psychological Narratives: The Social Basis of Understanding Reasons', The MIT Press, Cambridge, MA.
Jackendoff, R. (1983), 'Semantics and Cognition', The MIT Press, Cambridge, MA.
Jakobson, R. and Halle, M. (1956), 'Fundamentals Of Language', The Hague: Mouton.
Jameson, K.A. (2005), 'Culture and Cognition: What is Universal about the Representation of Color Experience?', *The Journal of Cognition & Culture* 5 (3-4), pp. 293-347.
Jameson, K. A. and D'Andrade, R. G. (1997), 'It's not really red, green, yellow, blue: an inquiry into perceptual color space', in: 'Color Categories In Thought And Language', eds. C. L. Hardin and L. Maffi, Cambridge University Press.
Juarrero, A. (1999), 'Dynamics in action : intentional behavior as a complex system', The MIT Press, Cambridge, MA.
Judd, D. B., MacAdam, D. L. and Wyszecki, G. (1964), 'Spectral Distribution of Typical Daylight as a Function of Correlated Color Temperature', Journal of the Optical Society of America, vol. 54, no. 8 (august 1964)
Kay, P. (1975), 'Synchronic Variability And Diachronic Change In Basic Color Terms', *Language In Society*, 4: 257-270.
Kay, P. and McDaniel, C.K. (1978), 'The Linguistic Significance Of The Meanings Of Basic Color Terms', *Language*, 54: 610-646.
Kelso, J.A. Scott (1995), 'Dynamic Patterns - The Self-Organization of Brain and Behaviour', The MIT Press, Cambridge, MA.
Kim, J. (1998), 'Mind in a Physical World.', MIT Press, Cambridge, MA.
Kwong, J.M.C. (2006), 'Why Concepts Can't be Theories', *Philosophical Explorations*, Vol. 9, Nr. 3, pp. 309-325.
Lakoff, G. (1987), 'Women, Fire and Dangerous Things', University of Chicago Press.
Lakoff, G. and Johnson, M. (1980, 2003 [revised]), 'Metaphors We Live By', University Of Chicago Press, Chicago.
Lakoff, G. and Johnson, M. (1999), 'Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought', HarperCollins Publishers.

Laurence, S. and Margolis, E. (1999), 'Concept and Cognitive Science', in: 'Concepts: Core Readings' (1999), The MIT Press, Cambridge, MA / London, England.

Latash, M. (2001), 'Mirror Writing: Adults Making A-non-B Errors?', open peer commentary on: E. Thelen, G. Schöner, C. Scheier and L.B. Smith (2001), 'The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching', *Behavioral and Brain Sciences* 24, pp.1-86.

Lee, T.W.; Wachtler, T. and Sejnowski, T. J. (2002) - 'Color opponency is an efficient representation of spectral properties in natural scenes', Vision Research 42 (2002), 2095-2103.

Lewis, D. (1972), 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50, pp 1249-58)

Locke, J. (1690), 'An Essay Concerning Human Understanding', P.H. Nidditch (ed.), Oxford University Press, Oxford, 1979.

Lucy, J. A. (1992), 'Language, Diversity and Thought - A Reformulation of the Linguistic Relativity Hypothesis', *Studies In The Social And Cultural Foundations Of Language* No. 12, Cambridge University Press.

Lucy, J. A. (1997a), 'Linguistic Relativity', *Annual Review Of Anthropology*, 26: 291-312.

Maclaury, R. (1997), 'Color and Cognition in Mesoamerica: Constructing Categories as Vantages', University of Texas Press, Austin.

Maclaury, R. (2002), 'Introducing Vantage Theory', *Language Sciences* 24: 493-536.

Maloney, L. T. (1992), 'A Mathematical Framework for Biological Color Vision', commentary on Thompson, E., Palacios, A. and Varela, F.J. (1992) – 'Ways of Coloring: Comparative Vision as a Case Study for Cognitive Science', *Behavioral And Brain Sciences* 15: 1-74.

Maloney, L. T., and Wandell, Brian A. (1986), 'Color constancy: a method for recovering surface spectral reflectance', *J. Opt. Soc. Am. A* 29, Vol. 3, No. 1 (1986)

Mandler, J.M. (2007), 'On the Origins of the Conceptual System', *American Psychologist* vol. 62, issue 8, pp. 741-751

Maravita, A. and Iriki, A. (2004), 'Tools for the body (schema)', *TRENDS in Cognitive Sciences* vol. 8, nr. 2.

Marchionni, C. (2008), 'Explanatory Pluralism and Complementarity: From Autonomy to Integration', *Philosophy of the Social Sciences*, vol. 38, pp. 314-333.

Markman, A.B. (2001), 'Are Dynamical Systems the Answer?', open peer commentary on: E. Thelen, G. Schöner, C. Scheier and L.B. Smith (2001), 'The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching', *Behavioral and Brain Sciences* 24, pp.1-86.

Markman, A.B. and Stillwell, C.H. (2004), 'Concepts á la Modal: An Extended Review of Prinz's Furnishing the Mind', *Philosophical Psychology*, vol. 17, no. 3., pp. 391-401.

Marr, D. (1982), 'Vision : a computational investigation into the human representation and processing of visual information', W.H. Freeman, San Francisco.

Maturana, H.R. and Varela, F.G. (1972), 'De máquinas y seres vivos', Editorial Universitaria, Santiago. English version: 'Autopoiesis: the organization of the living', in Maturana, H. R., and Varela, F. G. (1980), 'Autopoiesis and Cognition', Reidel, Dordrecht.

Maund, B. (1981), 'Colour - A Case for Conceptual Fission', *Australasian Journal of Philosophy*, Vol. 59, No. 3.

Maund, B. (1995), 'Colours: Their Nature and Representation', Cambridge University Press, Cambridge.

Mayr, E. (1961), 'Cause and Effect in Biology', *Science* 134: 1501-1506.

McCulloch, G. (2003), 'The life of the mind: an essay on phenomenological externalism', Routledge, London.

Menary, R. (2006), 'Introduction: What is Radical Enactivism?', in: Richard Menary (ed.), 'Radical Enactivism: Intentionality, Phenomenology and Narrative - Focus on the Philosophy of Daniel D. Hutto', John Benjamins Publishing, Amsterdam.

Merriam-Webster online dictionary [URL: http://www.m-w.com]

Millikan, R. G. (1984), 'Language, Thought and Other Biological Categories', The MIT Press, Cambridge, MA.

Millikan, R.G. (1996), 'Pushmi-pullyu representations', in: *Philosophical Perspectives* vol. IX (J. Tomberlin, ed.), pp.185-200, Atascadero Ridgeview Publishing. *Reprinted* in Mind and Morals (L. May and M. Friedman, eds.), pp145-161, The MIT Press, Cambridge, MA

Mithen, S. (1996), 'The Prehistory of the Mind: The Cognitive Origins of Art and Science', Thames and Hudson, London.

Nagel, T. (1974), 'What Is it Like to Be a Bat?', *Philosophical Review* LXXXIII, issue 4, pp. 435-50.

Nassau, K. (2001), 'The Physics and Chemistry of Color, 2nd Edition', John Wiley & Sons, Inc., New York.

Noë, A. (2004), 'Action in Perception', The MIT Press, Cambridge, MA.

Norman, D.A. (1999), 'Affordances, Conventions and Design', *Interactions* 6(3), pp. 38-43

O'Regan, K. and Noë, A. (2001), 'A sensorimotor account of vision and visual consciousness', Behavioral and Brain Sciences 24 (5), pp. 883-917.

Osherson, D.N. and Smith, E.E. (1981), 'On the adequacy of prototype theory as a theory of concepts', *Cognition* 9, pp. 35-58.

Peacocke, C. (1992), 'A Study of Concepts', The MIT Press, Cambridge, MA.

Pelphrey, K.A. and Reznick, J.S. (2001), 'Clothing a Model of Embodiment', open peer commentary on: E. Thelen, G. Schöner, C. Scheier and L.B. Smith (2001), 'The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching', *Behavioral and Brain Sciences* 24, pp.1-86.

Peschl, M. and Riegler, A. (1999), 'Does Representation Need Reality? Rethinking Epistemological Issues in the Light of Recent Developments and Concepts in Cognitive Science', in:

Understanding Representation in the Cognitive Sciences', Riegler, A., Peschl, M. and Von Stein, A. (eds.), Kluwer Academic/Plenum Publishers, New York.

Plato (360 B.C.), 'Pheadrus', translated by Benjamin Jowett; URL: http://classics.mit.edu/Plato/phaedrus.html, retrieved on February 23rd, 2009.

Poincaré, H. (1906), 'Les mathématiques et la Logique', *Revue de Métaphysique et de Morale'*, vol. 14, pg. 294-317.

Port, R. and Van Gelder, T. (eds.) (1995), 'Mind as Motion: Explorations in the Dynamics of Cognition' , The MIT Press, Cambridge, MA.

Prinz, J.J. (2002), 'Furnishing the Mind: Concepts and their Perceptual Basis', The MIT Press, Cambridge, MA.

Prinz, J. J. and Barsalou, L. W. (2000), 'Steering a course for embodied representation', in: Dietrich, E. and Markman, A.B. (Eds.), 'Cognitive dynamics: Conceptual and representational change in humans and machines', Lawrence Erlbaum Associates, Mahwah, NJ.

Putnam, H. (1975). 'Mind, language, and reality', Cambridge University Press, Cambridge MA.

Putnam, H. (1988), 'Representation and Reality'. The MIT Press, Cambridge, MA.

Quine, W. V. O. (1969), 'Natural Kinds'. in *Ontological Reality and Other Essays*: Columbia University Press.

Reeves, A. (1992), 'Areas of Ignorance and Confusion in Color Science', commentary on Thompson, E., Palacios, A. and Varela, F.J. (1992) – 'Ways of Coloring: Comparative Vision as a Case Study for Cognitive Science', *Behavioral And Brain Sciences* 15: 1-74.

Rey, G. (1994), 'Concepts', in: Guttenplan, S. (ed.) (1994), 'A Companion to the Philosophy of Mind', Blackwell Publishers, Oxford/Malden, MA.

Rizzolatti, G.; Fadiga, L.; Gallese, V. and Fogassi, L. (1996), 'Premotor cortex and the recognition of motor actions', *Cognitive Brain Research* 3, pp. 131-141.

Roberson, D., Davidoff, J. and Davies, I. (2000), 'Color Categories Are Not Universal: Replications and New Evidence From a Stone-Age Culture', *Journal Of Experimental Psychology: General*, 129: 369-398.

Rosch-Heider, E. (1972), 'Universals In Colour Naming And Memory', *Journal Of Experimental Psychology*, 93: 10-20.

Rosch, E. (1978), 'Principles of Categorization', in: Rosch, E. & Lloyd, B.B. (eds) (1978), 'Cognition and Categorization', Lawrence Erlbaum Associates Publishers, Hillsdale.

Rosen, R. (1991), 'Life Itself', Columbia University Press, New York.

Ryle, G. (1949), 'The Concept of Mind', University of Chicago Press, Chicago.

Rynasiewicz, R. (1996), 'Absolute Versus Relational Space-Time: An Outmoded Debate', *The Journal Of Philosophy*, #93, pp. 279-306.

Saunders, B. A. C. (1992), 'The Invention Of Basic Colour Terms', PhD Thesis, Utrecht University.

Saunders, B. A. C. and Van Brakel, J. (1997), 'Are there nontrivial constraints on colour categorization?', *Behavioral And Brain Sciences*, 20, pp. 167-228.

Shepard, R.N. (1987), 'Evolution of a Mesh between Principles of the Mind and Regularities of the World' (1987), in 'The Latest on the Best: Essays on Evolution and Optimality' (1987), John Dupré, ed., MIT Press, Cambridge.

Shepard, R.N. (1992): 'The Perceptual Organization of Colors: An Adaptation to Regularities of the Terrestrial World?', reprinted in 'Readings on Color, Volume 2 - The Science of Color' (1997), Alex Byrne + David Hilbert (eds.), MIT Press.

Shepard, R.N. (2001), 'Perceptual-Cognitive Universals as Reflections of the World' , *Behavioral and Brain Sciences* vol. 24 issue 3 (2001); reprinted from *Psychonomic Bulletin & Review*, 1994, 1, 2-28

Sklar, L. (1977), 'Space, Time and Spacetime', University of California Press, Berkeley.

Slors, M.V.P. (2001), 'The Diachronic Mind: An Essay on Personal Identity, Psychological Continuity and the Mind-Body Problem', Kluwer Academic Publishers, Dordrecht.

Sophian, C. (2001), 'Does Cognitive Development move Beyond Sensorimotor Intelligence?', open peer commentary on: E. Thelen, G. Schöner, C. Scheier and L.B. Smith (2001), 'The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching', *Behavioral and Brain Sciences* 24, pp.1-86.

Stern, D.N. (1985), 'The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology', Basic Books, New York.

Thelen, E. (1995), 'Time-Scale Dynamics and the Development of an Embodied Cognition', in Port, R. and Van Gelder, T., (eds.) (1995), 'Mind as Motion: Explorations in the Dynamics of Cognition', The MIT Press, Cambridge, MA.

Thelen, E., Schöner, G., Scheier, C. and Smith, L. B. (2001), 'The Dynamics of Embodiment: A Field Theory of Infant Perseverative Reaching', *Behavioral and Brain Sciences* 24, pp. 1-86.

Thelen, E. and Smith, L.B. (1994), 'A Dynamic Systems Approach to the Development of Cognition and Action', The MIT Press, Cambridge, MA.

Thompson, E. (1995a), 'Colour Vision - A Study In Cognitive Science And The Philosophy Of Perception', London / New York: Routledge.

Thompson, E. (1995b), 'Colour Vision, Evolution, and Perceptual Content', *Synthese* # 104.

Thompson, E. (2000), 'Comparative Color Vision: Quality Space And Visual Ecology', from 'Color Perception: Philosophical, Psychological, Artistic and Computational Perspectives' (pg.163-186), Vancouver Studies in Cognitive Science, Volume 9. Steven Davis (ed.), Oxford: Oxford University Press.

Thompson, E. (2006), 'Neurophenomenology and contemplative experience', in Philip Clayton (ed.), 'The Oxford Handbook of Science and Religion'. Oxford University Press.

Thompson, E. (2007), 'Mind in Life: Biology, Phenomenology and the Sciences of Mind', Harvard University Press.

Thompson, E., Palacios, A. and Varela, F.J. (1992), 'Ways of Coloring: Comparative Vision as a Case Study for Cognitive Science', *Behavioral And Brain Sciences* 15: 1-74.

Tononi, G. (2004), 'An Information Integration Theory of Consciousness', *BMC Neuroscience*, 5: 42.

Turvey, M.T. (1992). Affordances and Prospective Control: An outline of the ontology. *Ecological Psychology*, Vol. 4, pg. 173-187.

Turvey, M.T., Shaw, R.E., Reed, E.S. and Mace, W.M. (1981), 'Ecological laws of Perceiving and Acting: in Reply to Fodor and Pylyshyn (1981)', *Cognition* 9: 237-304

Van Gelder, T. (1998), 'The Dynamical Hypothesis in Cognitive Science ', *Behavioral and Brain Sciences* 21, pp. 615 –665.

Van Gelder, T. (1999), 'Revisiting the Dynamical Hypothesis ', Preprint No.2/99, University of Melbourne, Department of Philosophy. (URL:http://www.arts.unimelb.edu.au/~tgelder/papers/Brazil.pdf)

Van Gelder, T. and Port, R.F. (1995), 'It 's About Time: An Overview of the Dynamical Approach to Cognition', in R. Port and T. van Gelder (eds.), 'Mind as Motion: Explorations in the Dynamics of Cognition', The MIT Press, Cambridge, MA.

Van Hateren, J. H., and Van Der Schaaf, A. (1998), 'Independent component filters of natural images compared with simple cells in primary visual cortex', *Proceedings of the Royal Society of London B* 265: pp. 359-366.

Van Leeuwen, M. (2005), 'Questions For The Dynamicist:The Use of Dynamical Systems Theory in the Philosophy of Cognition', *Minds and Machines* 15, pp. 271-333.

Van Rooij, I., Bongers, R.M. and Haselager, W.F.G. (2002), 'A Non-representational Approach to Imagined Action ', Cognitive Science 26, pp.345-375.

Varela, F. J., Thompson, E. and Rosch, E. (1991), 'The Embodied Mind: Cognitive Science and Human Experience', The MIT Press, Cambridge, MA.

Vera, A.H. and Simon, H.A. (1993), 'Situated Action: A Symbolic Interpretation', *Cognitive Scienc*e, Vol. 17, pp. 7-48.

Wachtler, T., Lee, T.W. and Sejnowski, T. J. (2001) - 'Chromatic structure of natural scenes', Journal of the Optical Society of America A, Vol. 18 (2001), No. 1.

Wandell, B. (1989), 'Color Constancy and the Natural Image', reprinted in 'Readings on Color, Volume 2 - The Science of Color' (1997), Alex Byrne + David Hilbert (eds.), MIT Press.

Wiener, N. (1961), 'Cybernetics, or Control and Communication in the Animal and the Machine - 3 ed.', MIT and Wiley, New York/London.

Wouters, A. G. (2003), 'Four Notions of Biological Function', *Studies in History and Philosophy of Biology and Biomedical Science* 34: 633-668.

Wouters, A. G. (2004) - 'The Functional Perspective Of Organismal Biology', "Current Themes in Dutch Theoretical Biology", Lia Hemerick & Thomas A.C. Reydon (eds.), Dordrecht: Kluwer, 2004.

Ziemke, T. (2003). What's that thing called embodiment? In: *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Lawrence Erlbaum.

[Nederlandstalige Samenvatting (Summary in Dutch)]

-(Hoofdstuk 1)

De klassieke, 'Cartesiaanse' manier van denken over cognitie - waarin wordt gezegd dat de menselijke geest eigenschappen heeft die fundamenteel anders zijn dan de mogelijke eigenschappen van fysieke objecten - wordt gaandeweg minder populair onder filosofen en psychologen. Een belangrijke verwante theoretische positie is 'cognitivisme', welke onder andere inhoudt dat cognitie gedefinieerd dient te worden in termen van *interne*, vaak representationele processen.

En alternatieve manier om te denken over 'denken' is *belichaamde, gesitueerde cognitie*, waarin wordt gesteld dat een mentale toestand gedefinieerd dient te worden in relatie tot vele buiten het mentale liggende kenmerken - eigenschappen van het lichaam of de handelingsopties die het individu geboden worden door zijn omgeving in het bijzonder. Er zijn verschillende zienswijzen die, in verschillende samenstellingen, in de literatuur aangetroffen kunnen worden als passend onder de brede 'belichaamde, gesitueerde cognitie'-paraplu:

**Belichaamd* (cognitie heeft te maken met of wordt deels geïnstantieerd door lichamelijke processen; gebruikte notatie: $E_{(B)}C$);
**Gesitueerd* (cognitie betreft interageren met de omgeving; gebruikte notatie: $E_{(S)}C$);
**Enactief* (cognitie is een actief, dynamisch, gedragsgebaseerd proces; gebruikte notatie: $E_{(A)}C$);
**Uitgebreid* (processen uit de ongeving van het individu maken deel uit van het cognitie-proces; gebruikte notatie: $E_{(X)}C$);
**Geëncultureerd* (cognitie steunt deels op sociaal-culturele processen; gebruikte notatie: $E_{(C)}C$)

Algemene notatie voor niet nader gespecificeerde theorieën van het 'belichaamde, gesitueerde, enzovoort' soort: $E_{(i)}C$

Veel theorieën over cognitie bevatten, als een belangrijke component, een theorie over *concepten*. Het is de bedoeling in dit boek een theorie over concepten te ontwikkelen die past bij theorieën over belichaamde, gesitueerde, enactieve, uitgebreide en/of geëncultureerde cognitie. De centrale vraag van dit boek is dan ook als volgt: **Hoe kunnen we het concept 'concept' begrijpen op een manier die compatibel is met $E_{(i)}C$?** Dit is een filosofisch interessante vraag omdat 'concepten' als bouwstenen van gedachten op een verhoudingsgewijs eenvoudige manier in te passen zijn in een cognitivistisch verhaal over cognitie - bijvoorbeeld, als gedachten symbolische structuren zijn, kunnen concepten gezien worden als meer basale versies van dat soort symbolen -, maar het is niet direct duidelijk hoe concepten gedefiniëerd dienen te worden binnen $E_{(i)}C$.

De theorie zoals die ontwikkeld wordt in dit boek is vooral enactief ($E_{(A)}C$) van karakter. Een belangrijk probleem is dan dat veel $E_{(A)}C$- theorieën redelijk succesvol zijn in het verklaren van sensorimotor interacties van organisme en omgeving, maar minder goed toepasbaar zijn op gevallen waarin er daadwerkelijk sprake is van 'denken' op de manier zoals die term in de alledaagse taal gebruikt wordt - juist het domein waar veel cognitivistische theorieën (en de bijbehorende ideeën over concepten) het meest effectief zijn. Een belangrijk doel van dit boek is dit gemis verhelpen.

De globale structuur van de rest van het boek is als volgt: in hoofdstuk 2 worden enkele standaardtheorieën van concepten behandeld, alsmede hun zwakke plekken. Hoofdstuk 3 introduceert een nuttige $E_{(A)}C$-theorie waarmee basaal cognitie-gestuurd gedrag beschreven kan worden: het dynamisch bewegings-planningsveld van Thelen, Schöner, Scheier en Smith (2001). Hoofdstuk 4 gaat over kleurwaarneming, omdat er voor dat verschijnsel in de loop der tijd verschillende modellen ontwikkeld zijn die een verband suggereren tussen de basale sensorimotor disposities van een organisme (bijvoorbeeld de wijze waarop de eigenschappen van zijn netvlies en de daarop volgende neurale verwerking het waarnemen van kleuronderscheid mogelijk maken) en meer geavanceerd kleurgerelateerd gedrag (bijvoorbeeld aangaande de cultuurspecifieke betekenis van bepaalde kleurtinten, of zelfs wetenschappelijke concepten van wat 'kleur' nu eigenlijk is). Een doel van dit boek zal zijn dit verband - dus tussen basale sensorimotor processen enerzijds en hoger ontwikkeld kleurgerelateerd gedrag anderzijds - aan te passen op een zodanige manier dat er een werkbare, voor $E_{(i)}C$ geschikte theorie van concepten in het algemeen ontstaat.

-(Hoofdstuk 2)

Het blijkt bijzonder lastig te zijn een goede definitie te geven van wat een concept nu eigenlijk is: kandidaten zijn, onder andere, mentale representaties en vaardigheden. De $E_{(A)}C$-insteek impliceert die laatste optie: het omschrijven van het hebben van concepten in termen van het hebben van bepaalde vaardigheden.

De standaardtheorieën over concepten die worden besproken in dit hoofdstuk zijn:

*de *Klassieke Theorie* (een concept is een representatie die noodzakelijke en voldoende eigenschappen van een object codeert);
**Prototypetheorie* (een concept is een representatie die eigenschappen van een object op een gerangschikte manier codeert);
*de *Theorie-Theorie* (conceptuele structuur wordt bepaald door een mentale theorie).

Elk van deze theorieën heeft niet-triviale nadelen. Naast het kunnen oplossen van dit soort problemen zal een goede theorie van concepten moeten voldoen aan een serie eisen: een dergelijke theorie zal voldoende

verklarende *reikwijdte* moeten hebben, zal moeten kunnen verklaren hoe concepten *intentionele inhoud* en *cognitieve inhoud* kunnen hebben, hoe *conceptverwerving* plaatsvindt, hoe concepten *categorisatie* mogelijk maken, hoe eenvoudige concepten tot meer complexe concepten geombineerd kunnen worden (*compositionaliteit*), en hoe verschillende mensen in staat kunnen zijn hetzelfde concept te hebben (*publiciteit*). In het laatste hoofdstuk, hoofdstuk 10, wordt uitgelegd hoe de theorie, ontwikkeld in dit boek, aan deze eisen voldoet.

-(Hoofdstuk 3)

Thelen et al. (2001) hebben, gebruikmakend van dynamische systeemtheorie, een dynamisch bewegingsplanningsveld ontwikkeld, een model dat activiteit laat zien die congruent is met de fundamentele gedragskeuzes van jonge kinderen als ze de 'A-niet-B-fout' maken. Dit model is enactief van karakter, en maakt dus geen gebruik van interne representaties om dit cognitieve gedrag te verklaren. Eén van de belangrijkste tekortkomingen van dit model is dat het weinig aandacht besteedt aan meer geavanceerde cognitie: het model biedt een globale beschrijving van $E_{(i)}C$ gedrag in termen van een veldmodel; een belangrijke taak die in dit boek uitgevoerd wordt is het uitbreiden en aanpassen van dit model op zo'n manier dat conceptgerelateerd gedrag in het algemeen beschreven kan worden - dit zal voornamelijk gebeuren in hoofdstuk 8. Op deze manier kan er eveneens een antwoord gevonden worden op *Newton's vloek*: de neiging tot het negeren van contextuele effecten in het extrapoleren van kwantitatieve methodologie naar kwalitatieve ontologie. Het 'Radicality Manifold'-model dat zal worden ontwikkeld in dit boek zal een uitgebreider beschouwing bieden van de interactie van organisme, fysieke omgeving, sociale omgeving en concepten, als beschrijving van gedrag.

In in de volgende hoofdstukken wordt het model van Thelen en collega's in stadia uitgebreid: de eerste aanwijzing over het precieze karakter van sensorimotor gesitueerdheid (de basale interactie van individu met zijn omgeving) en sociale gesitueerdheid volgen in hoofdstuk 4, waar deze interactievormen worden bediscussieerd voor zover ze betrekking hebben op kleurwaarneming. Het idee hierachter is dat verschillende theorieën over kleurwaarneming veel informatie verschaffen over de manier waarop sensorimotor disposities een bijdrage leveren aan de eigenschappen van complexer gedrag. Deze ideeën zullen leiden tot een complexere, hoger-dimensionale versie van het veld van Thelen et al.; de eigenschappen van deze modelmatige ruimte, zoals deze geëxtrapoleerd worden uit het kleurvoorbeeld, volgen in hoofdstukken 5 en 6.

-(Hoofdstuk 4)

In de meest gangbare theorie over kleurwaarneming is er sprake van een fenomenale structuur die in drie dimensies is ingedeeld (helderheid, verzadiging en tint). Deze structuur is afgeleid van gedragsresponsen als

reactie op chromatische stimuli afkomstig uit de omgeving, en gevormd door de eigenschappen van het netvlies en de daaropvolgende neurale verwerking van die stimuli. Deze neurale verwerkingsmechanismen en de daarop gebaseerde perceptuele categorisatie zijn volgens voorstanders van deze theorie universeel, ondanks dat niet alle talen hetzelfde aantal basiskleurtermen hebben.

Critici van deze gangbare, universalistisch georiënteerde theorie, vaak voorstanders van linguïstisch relativisme, stellen dat bovenstaand verhaal zich schuldig maakt aan *decontextualisatie*: in sommige 'primitieve' culturen zijn de basiskleurwoorden die kleurcategorieën benoemen geen neutrale labels, maar woorden die veel meer semantische inhoud bezitten. Dat betekent dat kleurgerelateerd gedrag niet slechts verklaard kan worden op basis van een neuraal verwerkingsmechanisme, maar ook een hele belangrijke socioculturele component kent.

In dit boek wordt daarom een tussenpositie ontwikkeld, die het belang erkent van *zowel* een soortspecifieke, neurofysiologisch gefundeerde dispositie tot bepaalde kleurcategorisaties, *als* socioculturele invloeden. Dit betekent dat er een verband zou moeten bestaan tussen die basale sensorimotor disposities en hogere-orde socioculturele regelmatigheden. Met andere woorden: om recht te kunnen doen aan kleurgerelateerd gedrag zal de eerdergenoemde driedimensionale fenomenale structuur (die gedragsresponsen beschrijft zoals die gevormd worden door neurofysiologische eigenschappen) uitgebreid moeten worden om complexer gedrag (zoals gevormd door socio-culturele eigenschappen) te kunnen omvatten.

Dit idee, gekoppeld aan het dynamische bewegingsplannings*veld* als beschrijving van basaal enactief gedrag uit hoofdstuk 3, biedt een eerste glimp van een complexer gedrags*ruimte*-model. In de komende hoofdstukken zal een uitgebreidere beschrijving gegeven worden van dit model, en zal het veranderd worden in een *conceptuele* ruimte.

Er is echter een derde klasse eigenschappen, naast de neurofysiologische (het lichaam) en socio-culturele (sociale omgeving) eigenschappen die hierboven genoemd worden: *fysieke* omgevingseigenschappen. In hoofdstuk 4 wordt beschreven hoe een enactieve kleurwaarnemingstheorie (van Evan Thompson) aangevuld kan worden met een ecologische theorie van kleurwaarneming die het idee verdedigt dat er een congruentie bestaat van netvlieseigenschappen en eigenschappen van de omgeving, namelijk de chromatische structuur van het omgevingslicht (de theorie van Roger Shepard). Gecombineerd bieden deze theorieën een verhaal over kleurwaarneming dat stelt dat er een wederzijdse afstemming is van organisme en fysieke omgeving, waardoor de perceptuele en gedragsmatige mogelijkheden van dat organisme bepaald worden. Dit idee van interactie van organisme en fysieke omgeving kan gecombineerd worden met de eerder ontwikkelde ideeën over de interactie van en organisme met zijn sociale omgeving (waarbij beide in belangrijke mate

bepaald worden door zijn belichaamde eigenschappen), om collectief een redelijk complete omschrijving te beiden van het $E_{(i)}C$-gerelateerde gedrag van het organisme.

Samenvattend: de enactieve opvatting van 'gesitueerdheid' impliceert zelf al de relevantie van sociale handelingsmogelijkheden (als complement van fysieke/ecologische handelingsmogelijkheden) bij het verklaren van het gedrag van een organisme; de gebruikelijke universalistische theorie over de linguïstische aspecten van 'kleurcognitie' schetst een gedecontextualiseerd beeld van dat verschijnsel, dus een zekere mate van relativisme (socioculturele factoren die de lichaamsgebaseerde eigenschappen van het organisme aanvullen) is gewenst. Deze beide benaderingen gecombineerd bieden een min of meer compleet beeld van de organisme-omgeving-interactiedynamiek. In hoofdstuk 5 wordt nog het één en ander gezegd over de eigenschappen van de eerdergenoemde fenomenale kleurruimte, waarna in hoofdstuk 6 de conceptuele ruimte geïntroduceerd kan worden.

-(Hoofdstuk 5)

In dit hoofdstuk wordt, voortbouwend op ideeën van Kimberly Jameson, meer detail verschaft over de progressieve fragmentatie van de fenomenale kleurruimte. Beginnend met de meest eenvoudige categorisatie (donker vs. licht), bepalen de neurofysiologisch bepaalde kleurtintgevoeligheden interagerend met ecologische en socio-culturele factoren welke opeenvolging van kleurruimtesegmentaties het meest informatief is. Dus: de manier waarop kleuren benoemd en geconceptualiseerd worden, wordt beïnvloed door ecologische/omgevingsgerelateerde *en* socio-culturele factoren (i.e. de rol die objecten van specifieke kleuren spelen in, respectievelijk, de ecologische niche en de socio-culturele praktijk van het organisme).

Het in dit hoofdstuk gegeven voorbeeld van de Hanunóo (Filippijnen) - hun kleurwoorden hebben complexe correlaties die in het Nederlands of Engels niet direct met kleur te maken hebben - suggereert dat de driedimensionale fenomenale kleurruimte uitgebreid zou moeten worden met semantische connecties (waarin een kleurconcept niet slechts een tint, maar bijvoorbeeld ook een bepaald object of ritueel aanwijst, waaraan uiteraard weer andere betekenissen gerelateerd zijn); dit is de eerste stap op weg naar een model van een daadwerkelijke conceptuele ruimte als een hoger-dimensionale versie van perceptuele ruimte.

Het 'Interpoint Distance Model' van Jameson maakt als zodanig de weg vrij voor een theorie over de segmentatie van conceptuele ruimte, waarin complexere structuren gekoppeld worden aan meer basale structuren via mechanismen als 'cross-dimensional mapping' (zie ook sectie 6.9, waarin dit idee in verband gebracht wordt met de theorie van George Lakoff over metaforische conceptontwikkeling), alsmede aan de interactie van de

lichamelijke eigenschappen van het organisme met de eigenschappen van zijn sociale en fysieke omgeving.

In hoofdstuk 6 zal een gedetailleerder beschrijving geboden worden van concepten als gedragsdisposities, dat past bij dit idee van een 'conceptuele ruimte'; in hoofdstuk 7 en verder zal er uitgelegd worden hoe conceptuele ruimte (C-ruimte) samenhangt met M-, S- en P-ruimte (respectievelijk: lichamelijke eigenschappen; socio-culturele omgevingseigenschappen en fysieke omgevingseigenschappen).

-(Hoofdstuk 6)

In dit hoofdstuk wordt het idee gepresenteerd dat het concept 'kleur' een complex concept is, dat toepassingen heeft in verschillende contexten die niet zondermeer gereduceerd kunnen worden tot een enkelvoudige definitie: het overkoepelende concept 'kleur', gekarakteriseerd als een 'superpositie', omvat meerdere elkaar uitsluitende subconcepten. Op basis van dit inzicht kan gezegd worden dat de inhoud van een concept afhankelijk is van de (gebruiks-)context. Een bij $E_{(i)}C$ passende conceptdefinitie luidt dan als volgt: een concept is een gestructureerde gedragsdispositie van een belichaamd en gesitueerd organisme, waarbij gedrag in dit geval eveneens cognitie en spraak omvat. 'Conceptuele ruimte' is dan een modelmatige uitdrukking van de structuur die bestaat in de inferentiële verbindingen tussen verschillende gedragspatronen die een individu kan vertonen als *verantwoording* van zijn gebruik van een bepaald concept: het vertonen van bepaald gedrag in bepaalde omstandigheden door een individu impliceert het hebben van een concept, en dit vermoeden kan door anderen getest worden (bijvoorbeeld door vragen te stellen). De inhoud van concepten wordt verschaft door narratieven, die opgebouwd zijn uit de ervaringen, herinneringen en ideeën van het individu.

Conceptual enslavement is het fenomeen dat zo'n narratief, die een concept inhoud verschaft, een bepaald zwaartepunt kan hebben - ervaringen enzovoort die een verhoudingsgewijs grote bijdrage leveren aan de betekenis van het betreffende concept, en die vaker dan gemiddeld genoemd zullen worden als het individu zijn conceptgebruik uitlegt. Het detailniveau ('*granularity*') waarop een concept gebruikt of uitgelegd wordt kan van situatie tot situatie verschillen, en kan ertoe bijdragen dat twee individuën er achter komen dat ze een concept gemeen hebben, ondanks dat de één wellicht tot een veel dieper begrip van dat concept in staat is. In de meeste alledaagse omstandigheden, echter, is het verkennen van een concept tot dergelijke dieptes niet nodig, en volstaat het erkennen van het conceptgebruik van de ander op een laag detailniveau. Dit belicht een belangrijk aspect van concepten: een concept is wat je doet/zegt/denkt in een bepaalde context, een praktijk die deels afhangt van de wijze waarop een soortgenoot jou als conceptgebruiker beoordeelt. Het hebben van concepten is, in ieder geval deels, het ontvangen van de bevestiging van anderen dat je die concepten op een acceptabele manier gebruikt (zie sectie 9.2 voor meer hierover). Deze ideeën tezamen resulteren in de

SuperpositieTheorie van Complexe Concepten (SToCC), waarbinnen complexe concepten begrepen worden als speciale gevallen van een algemene conceptentheorie.

Op basis van het idee van een conceptuele ruimte (als modelmatige uitdrukking van de concepten van een individu) als spectrum dat reikt van de meest basale sensorimotor vermogens tot complexe, abstracte ideeën, is het mogelijk de ontwikkeling van een conceptueel systeem te schetsen. In hoofdstuk 6 wordt een proces beschreven dat uit de volgende vier fasen bestaat:
[Fase 1]: sensorimotor bewegingsbegrip;
[Fase 2]: correlatie van sensorimotor kennis en linguïstische codering;
[Fase 3]: belichaamde en gesitueerde modaliteitstransformaties;
[Fase 4]: correlatie van belichaming en abstractie.

Het hoofdstuk bevat eveneens een beschrijving van de daaropvolgende ontwikkeling van conceptuele ruimte, gebaseerd op de suggesties uit hoofdstuk 5 (waar het gaat over de segmentatie van perceptuele ruimte). Veel gedetailleerde en soms zelfs nieuwe concepten kunnen ontstaan op basis van splitsing van conceptuele ruimte, waarin er sprake is van de segmentatie van conceptuele ruimte als een *meervoudig ingebed manifold*. Conceptuele ruimte begrijpen als een ingebed manifold betekent zeggen dat conceptuele ruimte opgebouwd is uit in elkaar passende regio's die subconcepten en de daartussen bestaande inferentiële verbindingen voorstellen, 'orthogonaal' daarop georganiseerd langs granulariteitsgradiënten. Sommige van de meest fundamentele onderverdelingen van conceptuele ruimte hangen af van de categorisaties die afgedwongen worden door de de eigenschappen van ons lichaam en onze zintuigen.

Dit hoofdstuk bevat een tussentijdse evaluatie van SToCC, waarin deze theorie vergeleken wordt met de prototypetheorie, theorie-theorie, de theorie van Jerry Fodor en 'Conceptual Role Semantics', en deze suggereert dat SToCC positief afsteekt tegen deze andere theorieën. Een belangrijk punt betreft het verschil tussen prototypes (uit de prototypetheorie) en 'enslavers': prototypes leggen gezamenlijk de definitie van een concept vast, terwijl een 'enslaver' nu juist inferentie richting ideeën en gedrag mogelijk maakt die als passend gezien kunnen worden voor iemand die meent een bepaald concept te hebben.

Het is echter duidelijk dat een gedetailleerdere beschrijving van de relaties tussen conceptuele ruimte ('C-ruimte') en lichamelijke eigenschappen, socio-culturele omgevingseigenschappen en fysieke omgevingseigenschappen (respectievelijk: M-, S- en P-ruimte) nodig is: hoe komt het dat een conceptuele ruimte een specifieke structuur heeft? Hoe kunnen we meer zeggen over de wijze waarop een belichaamd individu ingebed is in zijn omgeving, en welke weerslag is hiervan te zien in zijn concepten?

-(Hoofdstuk 7)

In hoofdstuk 7 wordt de eerste stap gezet op weg naar een antwoord op de hierboven genoemde vragen: er wordt een theorie over *representatie* ontwikkeld die past bij $E_{(i)}C$. Dit is een belangrijke stap omdat er in dit boek tot nu toe een belangrijke kloof zichtbaar is: tussen de enactieve ($E_{(A)}C$) benaderingen van Thelen et al.'s dynamische bewegingsplanningsveld, Thompson's kleurwaarnemingstheorie en de gedragsgeoriënteerde conceptdefinitie uit hoofdstuk 6 enerzijds, en de op interne representatie gebaseerde standaard concepttheorieën anderzijds. In dit hoofdstuk gebruik ik het 'Radicale Enactivisme' van Dan Hutto om een helderder beeld te krijgen van de rol van representatie in SToCC. Radicaal Enactivisme verdedigt het idee dat er geen sprake is van representationele inhoud (begrepen als 'dingen' die zich 'in het hoofd' bevinden), in ieder geval niet in beschrijvingen van meer basale vormen van enactie.

SToCC, hiermee contrasterend, stelt dat het mogelijk is een sluitend verhaal over representatie te ontwikkelen waarbinnen het mogelijk is de functie van een representatie te laten vervullen (namelijk: het instantiëren van een betekenisvolle relatie tussen individu en omgeving) zonder dat er sprake zou hoeven zijn van het ontologisch problematische 'verdinglijken' van interne mentale processen. Allereerst kan er gezegd worden dat Andy Clark, met zijn notie 'representation-hungry problems', laat zien dat er cognitieve taken zijn die soms de re-presentatie, in het geheugen bijvoorbeeld, nodig maken van objecten die niet op betrouwbare wijze aanwezig zijn in de onmiddelijke omgeving. Gebaseerd op distincties van Fred Dretske tussen verschillende soorten representatie kan SToCC stellen dat we interne toestanden kunnen hebben die geen representaties van het klassieke soort zijn (dus interne representaties van externe objecten), maar desalniettemin representaties zijn omdat ze een bijdrage leveren aan de dynamische organisme-omgeving-interactie vanwege een specifieke relatie met externe toestanden die ze uitdrukken; deze interactie betreft eveneens bijdragen van lichamelijke en omgevingsgerelateerde krachten ('enablings' ['mogelijkmakingen'] en 'constraints' ['inperkingen']). Deze 'representaties' zijn geen mentale entiteiten, maar mogelijkmakende of inperkende factoren in die organisme-omgeving interactiedynamiek.

Deze dynamiek betreft *wederzijdse* mogelijkmakingen en inperkingen: de lichamelijke eigenschappen van het individu, zijn sociale en fysieke omgeving in collectieve interactie realiseren een specifiek gedragsprofiel met nieuwe eigenschappen, waarvan de belangrijkste is de eigenschap conceptgerelateerd gedrag te vertonen: het individu zelf, belichaamd en ingebed in een bepaalde omgeving, heeft concepten. De metafysische structuur waarvan hier sprake is, is 'dynamical dimensioned realization'. De fluïditeit van deze dynamische interacties suggereert bovendien dat concepten niet onderhevig zijn aan rigide *waarheidsvoorwaarden*, maar aan *gebruikstoepasselijkheidsvoorwaarden*.

-(Hoofdstuk 8)

De wederzijdse gerelateerdheid van de lichamelijke eigenschappen van het individu, de sociale en fysieke eigenschappen van zijn omgeving en het conceptgerelateerde gedrag dat ontstaat uit deze dynamische interactie kan weergegeven worden in een model dat gebruik maakt van afzonderlijke maar gerelateerde ruimtes; dit model wordt uitgelegd in hoofdstuk 8. In dit model worden het dynamische bewegingsplanningsveld uit hoofdstuk 3, fenomenale kleurruimte uit hoofdstukken 4 en 5 en de conceptuele ruimte uit hoofdstuk 6 aangepast en samengevoegd tot een nieuw model, het 'Radicality Manifold' (RM).

Dit model beschrijft eigenschappen van de gedragsmatige ('B-ruimte'), biomechanische ('M-ruimte'), fysieke ('P-ruimte'), sociale ('S-ruimte') en conceptuele ('C-space') domeinen. Elk van deze domeinen kan uitgedrukt worden als een dispositionele ruimte, en de interactie van deze ruimtes wordt gedefinieerd in termen van 'affordances' (handelingsmogelijkheden): elk eigenschapstype roept mogelijkmakingen of juist inperkingen op voor andere eigenschapstypen.

Dit model laat zien hoe conceptuele vermogens ontstaan, via 'dynamical dimensioned realization', uit de interactie van fysieke, sociale en biomechanische processen. De beschikbaarheid van een conceptuele beschrijving maakt een *verklaring* mogelijk (als onderscheiden van slechts een *beschrijving*) van een specifiek gedragspatroon. Anders gezegd: de eigenschappen van de fysieke omgeving, sociale omgeving en lichamelijke biomechanica interageren op zo'n wijze dat er in het gedrag een structurele regelmatigheid ontstaat, namelijk een conceptuele dispositie. In dit hoofdstuk worden verschillende interactievormen tussen de ruimtes besproken: 'Semiogenetic Engine', 'Social Preconceptual Processing', 'Affordance-Effectivity Balance' en 'Distal Ecological Dynamics'.

-(Hoofdstuk 9)

Hoofdstuk 9 behandelt een aantal implicaties van het RM-model: het onstaan van betekenis en normativiteit, het inherent circulaire karakter van het model, de mogelijkheid binnen RM tot conceptindividuatie, en de epistemologische ideeën die in RM verstopt zitten.

Zeer basale lichamelijke syntax kan een bijdrage leveren aan het onstaan van een praktijk van participatoire betekenisgeving: het belichaamd interageren van individuën is een voorbeeld van het sociaal co-construeren van betekenisvolle interactieprofielen. Het kunnen herkennen van dat soort profielen is de basis van inhoud-toeschrijving, ons sociale spel waarin we zo handelen dat we voldoen aan de eisen van conceptbezit, zowel reflexief (jezelf interpreteren als handelend op zo'n manier dat het toeschrijven van concepten gerechtvaardigd is) en attributief (het hebben van eigenschappen en het vertonen van gedrag waardoor anderen geneigd zijn concepten aan dat individu toe te schrijven).

Het gebruiken van concepten in de uitleg van RM is een fundamenteel circulair proces, onder andere omdat iemand zichzelf interpreteert als conceptbezitter tegen een achtergrond van sociaal geaccepteerde omgangsvormen die door die persoon zelf in interactie met anderen uit zijn sociale niche geïnstantieerd worden. Een dergelijke circulariteit sluit goed aan bij een breder fenomeen: levende, autopoietische systemen moeten begrepen worden als systemen die zogenaamde *impredicatieve lussen* implementeren. RM als een binnen $E_{(i)}C$ passend model van conceptgerelateerd handelen vormt een uitdrukking hiervan.

Het individueren van concepten is in RM in principe mogelijk in termen van de verschillende mogelijkmakingen en inperkingen afkomstig van de ervaringen en ideeën zoals die vervat zijn in de narratieve jurisprudentie van een concept. Als we proberen het gedrag van anderen te verklaren kunnen we ze concepten toeschrijven, welke geïndividueerd kunnen worden door inferenties te plegen over de genoemde ervaringen en kennis op een grofkorrelig detailniveau.

Eén van de epistemologische consequenties van het RM-model is dat de verklarende focus nadrukkelijk op *het individu als geheel* gericht is: de verschillende deelverklaringen die gerelateerd zijn aan de afzonderlijke ruimtes van het RM-model worden opgesteld met expliciete aandacht voor het belichaamde, in zijn omgeving ingebedde individu. De deelverklaringen betreffen dus verschillende aspecten van één en dezelfde complexe organisme-omgeving-interactiedynamiek.

-(Hoofdstuk 10)

In het afsluitende hoofdstuk worden een aantal losse eindjes aan elkaar geknoopt. Een aanval op één van de belangrijke inspiratietheorieën van het RM-model, fenomenale kleurruimte, blijkt in het geval van RM minder effectief. De tegenwerping is dat de metrische structuur die door voorstanders in de perceptuele ruimte wordt aangebracht een empirische precisie veronderstelt die eenvoudigweg niet door introspectie (het 'meetinstrument' waarmee die structuur wordt vastgesteld) verschaft kan worden. De ruimtes uit het RM-model, echter, worden niet opgebouwd op basis van introspectie; bovendien wordt expliciet gesteld dat deze ruimtes nergens in het individu, zijn geest of waar dan ook aanwezig zijn, in tegenstelling tot de perceptuele ruimte in sommige interpretaties van die theorie - het gaat in het geval van RM om een model dat congruent geacht wordt te zijn met gedrag van een individu.

De theorie aangaande conceptuele ruimtes van Peter Gärdenfors vertoont enkele overeenkomsten met het RM-model, net als (maar dan op een andere manier) het concept-empirisme van Jesse Prinz. Elk van beide theorieën schiet echter tekort ten aanzien van $E_{(i)}C$-eisen, omdat ze - kort gezegd - representatie te klassiek-cognitivistisch opvatten: de specifieke interpretatie van representatie zoals die in hoofdstuk 7 ontwikkeld wordt (en

waarvan de implicaties in hoofdstukken 8 en 9 aan bod komen) is wél in staat die $E_{(i)}C$-eisen in te willigen.

In hoofdstuk 10 wordt RM toegepast op een concreet voorbeeld, namelijk conceptgebaseerd voor- en vroegschools onderwijs. Enkele van de aanbevelingen, op basis van RM, zijn: conceptgebaseerd onderwijs (als onderscheiden van onderwijs gericht op feitenkennis) kan kinderen helpen in het trainen van het contextafhankelijk interpreteren van gebeurtenissen, alsmede in het trainen van het begrip dat het kind van zijn eigen fysieke en sociale gesitueerdheid heeft; een multimodale presentatie van leermaterialen benadrukt de wederzijdse afhankelijkheid van veel begripsdefinities; en het hebben van concepten is in belangrijke mate een sociale eigenschap - in conceptgebaseerd onderwijs zou daarom extra aandacht besteed moeten worden aan het verantwoorden, door kinderen, van hun conceptgebruik, en aan het trainen van de analytische instelling die daarbij hoort.

Tenslotte wordt in hoofdstuk 10 samengevat hoe RM voldoet aan de eisen voor een conceptentheorie zoals die in hoofdstuk 2 werden geformuleerd.

[Acknowledgements]

Philosophy is about coming to grips with how the world works, and I believe that the most important fact one can come to understand is that very little happens without the support of many other people. The people listed below all contributed in some way to the process that resulted in this book, and I thank each and every one of them.

** Rob van der Sandt, Anna Bosman and Erik Myin for reading and judging the manuscript;

** Marc Slors, Tjeerd van de Laar, Derek Strijbos and Sander Voerman for many great discussions full of incisive, useful comments; additional thanks to Tjeerd for being a great guy to share an office with;

** My colleagues at the philosophy department of Radboud University Nijmegen, including Ton Derksen, Monica Meijsing, Chris Buskes and Arno Wouters, for support and feedback;

** The members of the 'Dynamical Systems Group', for many meetings filled with interesting, stimulating exchanges of ideas - our intermittent process of sharing has been an extremely important influence upon this book, and I hope you all like what I came up with;

** My friends and colleagues at the pedagogical sciences department of Radboud University Nijmegen, for a very pleasant and educational year of teaching: I learnt a lot, working with you;

** My friends and colleagues at the Academy for Leisure of the NHTV University Of Applied Sciences, for offering me an exciting new place to teach and do research;

** All the students that I've had the pleasure of interacting with - I've learnt much more from you than I could ever teach you;

** My family, for their unquestioning support;

** Joëlle Blankespoor, Yke Schippers, Carine Pots, Judith Rutte and Lars Eriksson for being good friends over the past few years;

** The good people at Universal Press in Veenendaal for helping me turn the virtual into the actual, and for their professional patience and support when I wanted something out of the ordinary for the cover;

** You, the reader, for taking an interest in what I have to say.

[About The Author]

Marco van Leeuwen (1976, Schiedam) studied philosophy at Leiden University and Erasmus University Rotterdam. He graduated in Leiden, having written a thesis on Daniel Dennett's philosophy of cognition, as well as a minor thesis on the relationship between philosophy and literature. From 2003 until 2007 he was a part of the philosophy department of Radboud University Nijmegen as a junior researcher, where he worked on his dissertation. Initially, his research was focused on the philosophy of colour perception (both the ecological and the anthropological aspects), but over time the focus of the project shifted, broadening into an investigation of some of the implications of adopting the embodied/embedded cognition paradigm in the philosophy of psychology. From 2007 until 2008, he taught at the pedogical sciences department of Radboud University. In 2008, he started teaching and doing research at the Academy for Leisure of the NHTV University of Applied Sciences (Breda, The Netherlands), where he is contributing to the creation of a new academic leisure sciences bachelor programme. He has published in refereed publications on dynamical systems theory, experience and new media, and his current research focuses on experience, meaning, normativity and multidisciplinarity as these themes pertain to leisure.